

## SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: Minnafield Examiner #: \_\_\_\_\_ Date: 3/6/02  
 Art Unit: 1645 Phone Number 30 53394 Serial Number: 91004395  
 Mail Box and Bldg/Room Location: CM1-8E12 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*  
 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Recombinant P37/FlaA as a Diagnostic Reagent  
 Inventors (please provide full names): Robert D. Gilmore Jr ; Barbara J.B. Johnson

Earliest Priority Filing Date: Jan 8, 1998

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

SEQ SEARCH → SEQ 1 (1-1665)  
 SEQ 2 (-341)  
 SEQ 3 (1-21)

Point of Contact  
 Mona Smith  
 Technical Information Specialist  
 CM1 6A01  
 Tel: 308-3278

RECEIVED  
 MAR-6-2002  
 (STIC)

Please search attached claims

Reviewed  
 3/26/02

Lyme Disease + Borrelia Burgdorferi

Diagnostic

Serodiagnostic antigen

outer sheath protein of periplasmic flagella

Thanks  
 Minnafield

## STAFF USE ONLY

## Type of Search

## Vendors and cost where applicable

Searcher: M. Smith NA Sequence (#) 2 STN \_\_\_\_\_  
 Searcher Phone #: \_\_\_\_\_ AA Sequence (#) 1 Dialog \_\_\_\_\_  
 Searcher Location: \_\_\_\_\_ Structure (#) \_\_\_\_\_ Questel/Orbit \_\_\_\_\_  
 Date Searcher Picked Up: \_\_\_\_\_ Bibliographic X Dr. Link \_\_\_\_\_  
 Date Completed: 3/25/02 Litigation \_\_\_\_\_ Lexis/Nexis \_\_\_\_\_  
 Searcher Prep & Review Time: 60 Fulltext \_\_\_\_\_ Sequence Systems \_\_\_\_\_  
 Clerical Prep Time: \_\_\_\_\_ Patent Family \_\_\_\_\_ WWW/Internet \_\_\_\_\_  
 Online Time: 60 Other \_\_\_\_\_ Other (specify) \_\_\_\_\_



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 23:36:44 ; Search time 2486.21 Seconds

(Without alignments)  
11034.792 Million cell updates/sec

Title: US-09-004-395-1

Perfect score: 1663  
Sequence: 1 atgataatcttttttcaaaa.....atatccttgaagaagaattt 1663

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_scs:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hgo\_hum:\*  
31: em\_hgo\_in:\*  
32: em\_hgo\_rod:\*  
33: em\_hgo\_hum:\*  
34: em\_hgo\_inv:\*  
35: em\_hgo\_rod:\*  
36: em\_hgo\_other:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1637	98.4	1655	1	BBU62900	U62900 Borrelia bu
2	1210.6	72.8	13271	1	AE001168	AE001168 Borrelia
3	435.8	26.2	10016	1	AE001167	AE001167 Borrelia
4	314	18.9	332	1	BBCHEAGEN	X91907 B. burgdorferi
5	164.4	9.9	2700	1	BBU28962	U28962 Borrelia bu
6	108.6	6.5	1140	6	A24079	A24079 T. hyodysent
7	108.6	6.5	1260	1	THFLAAL1	X63006 T. hyodysent
8	93.2	5.6	67970	3	PFMAL1P3	AL031746 Plasmodu
9	89.4	5.4	173957	2	AC006986	AC006986 Homo sapi
10	85.8	5.2	110000	2	PFMAL1P1_0	AL034557 Plasmodu
11	84.8	5.1	1128	3	MPU15677	U15677 Myrmecia pl
12	81.2	4.9	242513	2	AC079314	AC079314 Homo sapi
13	81	4.9	175229	2	CNSO1RG0	AL159141 Homo sapi
14	81	4.9	209157	9	CNSO1DM4	AL031632 Human chr
15	79.6	4.8	179537	2	AC013513	AC013513 Homo sapi
16	79	4.8	2192	3	AF298207	AF298207 Dictyoste
17	78.8	4.7	253305	3	PFMAL1P37	AL034559 Plasmodu
18	78.4	4.7	60604	2	AC023456	AC023456 Homo sapi
19	78	4.7	19124	6	PFAYARIA	L40608 Plasmodu
20	78	4.7	19124	6	AR089157	AR089157 Sequence
21	77.2	4.6	2718	1	AF250386	AF250386 Carsonell
22	77	4.6	256172	2	AC005139	AC005139 Plasmodu
23	77	4.6	310779	2	AC005140	AC005140 Plasmodu
24	76.4	4.6	713	3	AF034389	AF034389 Plasmodu
25	76.2	4.6	14867	3	AE001398	AE001398 Plasmodu
26	76.2	4.6	224448	2	PFMAL1P4	AL034477 Plasmodu
27	75.8	4.6	108908	3	PFMAL1P8	AL034560 Plasmodu
28	75.6	4.5	12029	3	AE001431	AE001431 Plasmodu
29	75.6	4.5	234112	3	PFMAL1P2	AL035475 Plasmodu
30	75.4	4.5	160891	9	AL512410	AL512410 Human DNA
31	75.2	4.5	86827	3	PFMAL1P5	AL034556 Plasmodu
32	75	4.5	1156	3	MPU15678	U15678 Myrmecia pl
33	74.8	4.5	222512	9	AC006335	AC006335 Homo sapi
34	74.6	4.5	186431	2	AC022281	AC022281 Homo sapi
35	74.6	4.5	213530	2	AC090493	AC090493 Mus muscu
36	74	4.4	9000	3	TETRRTRNA	L28677 Tetrahymena
37	74	4.4	14001	3	PFMAL1P3	X95276 P. falciparu
38	74	4.4	47296	3	AF160864	AF160864 Tetrahyme
39	74	4.4	66441	3	PFMAL1P4	AL031747 Plasmodu
40	74	4.4	293431	2	PFMAL1P4	AL049181 Plasmodu
41	73.8	4.4	53932	2	AC023371	AC023371 Homo sapi
42	73.8	4.4	208684	2	AC072048	AC072048 Mus muscu
43	73.6	4.4	178193	2	AC068949	AC068949 Homo sapi
44	73.6	4.4	186431	2	AC022281	AC022281 Homo sapi
45	73.2	4.4	130349	9	AC011593	AC011593 Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS BBU62900 1655 bp DNA BCT 15-JAN-1997  
DEFINITION Borrelia burgdorferi flagellar filament outsheath protein (flaA) gene, complete cds, and chemotaxis histidine kinase (cheA) gene, partial cds.  
ACCESSION U62900.1 GI:1575445  
VERSION 1  
KEYWORDS  
SOURCE  
ORGANISM Borrelia burgdorferi  
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.  
REFERENCE 1 (bases 1 to 1655)  
AUTHORS Ge,Y. and Charon,N.W.  
TITLE An unexpected flaA homolog is present and expressed in Borrelia burgdorferi  
JOURNAL J. Bacteriol. 179 (2), 552-556 (1997)  
MEDLINE 97144545  
REFERENCE 2 (bases 1 to 1655)





Db 1321 TGTATTAAGTTATGAGACTAGCGGAAGTGAATCCCTTGTAAATTAAGGACACAGA 1380  
Qy 1381 aacctttaaagagctttaaagcttagaagaaaaattctatcgctgaagctcttcca 1440  
Db 1381 AACTTTAAAGAGATTTTAAAGCTTAGAGAAAAATTTCTATCGCTGAAGCTCTTCCA 1440  
Qy 1441 aaacttctgaagaaagcttagagagtgaaaaaacctgaaagaaatcatctccgaaaaatagct 1500  
Db 1441 AACTCTTGTAAGAAAGATTGAGAGTGAAGAACTGAAGAAATCATCTCGAAAAATTAAGT 1500  
Qy 1501 ttaaatataatgtaagactaccctaaaggttgccttctterroyasnacataltaaaataa 1560  
Db 1501 TTAATTAATTAATGTAAGAGCTAACCTAAAAAGTTGCTTT-----ACATATTAAATAA 1562  
Qy 1561 taggaatagatataatggaataatagatttggaataatgaagagctttagagatttttt 1620  
Db 1553 TAGAAGTAATAGATATGCAAAATATTGATTGCAAAATGCAAGAGCTTTAGAGATTTTTT 1612  
Qy 1621 tgaagaaagctcaaaactctgtagatatccctgaaagagaatatt 1663  
Db 1613 TGAAGAAAGCTCAAAATCTTTAGATATCTTTGAAGAGATATTT 1655  
RESULT 2  
LOCUS AE001168 13271 bp DNA BCT 15-DEC-1997  
DEFINITION Borrelia burgdorferi (section 54 of 70) of the complete genome.  
ACCESSION AE001168 AE000783  
VERSION AE001168.1 GI:2688598  
KEYWORDS  
SOURCE Lyme disease spirochete.  
ORGANISM Borrelia burgdorferi  
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
burgdorferi group.  
1 (bases 1 to 13271)  
Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.,  
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
Gutin,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D.,  
Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,  
Salzberg,S., Hanson,M., Van-Vugt,R., Palmer,N., Adams,M.D.,  
Gocayne,J.D., Weldman,J., Uteerback,T., Wathey,L., McDonald,L.,  
Artlich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,  
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.  
Genomic sequence of a Lyme disease spirochete, Borrelia  
burgdorferi  
Nature 390 (6660), 580-586 (1997)  
JOURNAL MEDLINE 98065943  
REFERENCE 2 (bases 1 to 13271)  
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,  
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
Gutin,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D.,  
Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,  
Salzberg,S., Hanson,M., Van-Vugt,R., Palmer,N., Adams,M.D.,  
Gocayne,J.D., Weldman,J., Uteerback,T., Wathey,L., McDonald,L.,  
Artlich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,  
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.  
Direct Submission  
Submitted (12-DEC-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
TITLE JOURNAL Location/Qualifiers  
FEATURES source 1..13271  
gene /organism="Borrelia burgdorferi"  
/db\_xref="taxon:139"  
66..1100  
/gene="BB0668"  
66..1100  
/gene="BB0668"  
66..1100  
/note="similar to GP:1575447 percent identity: 98.83;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="flagellar filament outer layer protein (flaa)"  
/protein\_id="AAC67025.1"  
/db\_xref="GI:2688608"  
/translation="WYMKRKAISLIFLLSTVLPFAQETDGLAGSKRAPEGLVDF

AEIARDPSSTRLLDTNVVDYVYGASGIVPEDMVVDLGINNMSVLLTPSARLOAVYK  
NSYVAPAVYKSESKRACDTLLCYRVLPFSXSSASAMIMPFKIPFSGSSGNOFLK  
GLIDNITKMEIKVSVISLGEIDLEFDMGMEIATISMGLTKRGADILKMSNP  
YIPNISIRIIRKDVDPNYPPLASSKMRFAFVSKSHSEKQNFIFYVLDVLDKLSV  
SIDSIDSESVFVYETSGTESIRLKAHETFRVLKREKISMPEGSPONFVEKIES  
EKPEESPKN"  
1168..3762  
/gene="BB0669"  
1168..3762  
/note="similar to PID:1113815 SP:044737 GB:AE000783  
percent identity: 99.31; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis histidine kinase (cheA-2)"  
/protein\_id="AAC67024.1"  
/db\_xref="GI:2688607"  
/translation="METLDLENELGVPPEPAONLVILLENNIMSLDDPNNSPTID  
ETIRAATTLKSSASLDMLSDPTHEVEVFDAIRGKKNINNDLVLLSLSDYIK  
EMIALRIDGVVYLNDISDLKSLQFLVIDDQFFIKFSDGNSIKNNFCLSESDLEIR  
EGLIGQGVKIVSVFNSNSNEVNSGLKIFNLKMLGSLVHTIPYEQI1IEDPLK  
RVDYLYISDIEGVKSLDNLNIESYLVEDEPFVKEELKRLADEIDVLDNSNVP  
DNDFETDEISDLLLEVENOKLFKVRIDPFKDPNPMATISGLMLQALSLGKIFKSP  
DSSELADKRFEDFVIYLLISNTSEESIAKKNLPDVSHEIKVNLSELSKSVLAKED  
DEAPFKKNKIKKNSPISVLIRDSKKIDYILNLVSEANISKSYNOINSEMTIFY  
NENFYDQESFORNFDLIDKIVKDGTLLEDETHESINLSFKERKALKDSEIV  
NSFRLQNFKMTSGRLSRIITDLHESVLTMTPLINISFRTVRVDRSLKNTLV  
NLKMEGETELDKSVIDLVDPLMHCVANSMDQLEVEERVKSGSKACTIILRAN  
EGNVI5IEIDDGIGIDPKVIRKRLKGTIKEDAI5DEFLINLIPAPGFAVOYT  
DL5RGVGLVVKRSTKELNGTLTIESEIGTIFKIKLPLTVIIOGLVKGSEYI  
VPLNVL5ETHRTEHDIKLENHEYNIRDEYISLRLDKLFNTRDLSLIEKFLI  
VNTSNKIAIVD5SLIGEDFVVKPIKDKFSSAGIVGATTLGNGAVVLLIDFKLE  
DLQDKTE"  
3769..5169  
/gene="BB0670"  
3769..5169  
/gene="BB0670"  
/note="similar to GP:1881576 percent identity: 98.93;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="purine-binding chemotaxis protein (cheW-3)"  
/protein\_id="AAC67023.1"  
/db\_xref="GI:2688606"  
/translation="MQIKETIFGPKIIDDKNSKLTNDFKVVSEFGSDHYLVDM  
OYKEIRKSNMFTYVPAKRYVAGIDNIRGEIIPIDIRIMNLEFNKDKDIEDLVLM  
ENDLIGYIVKINNVE5IDSLLDDPVL5ODSLINIVIGVYDNKILINLVPI  
FNGEGERKLKDPONFVEKSDP5CDNLIIDLEKSDINQVYSSKDKGNNSTLNN  
TAAPNLENIRKNLKTSFNASLVNDVLEKGVKFNVDYDILPCDSFLNEPFSKSG  
NLNGADOLEEFKNEIVSRSLSPNNNL5INVL5EISGSGKETMALNAL5EYVKEP  
KLVAINDLSKVET5SRVSE5IG5EISY5RNS5FQ5G5V5K5FK5EIS5NVL5E5  
DAF5SDP5NL5G5V5K5F5L5C5F5DSKQ5DILINL5IA5K5IGALLV5GDNE5ELKNNDY  
F5KE5K5RY5N5L5RK5V"  
5181..5666  
/gene="BB0671"  
5181..5666  
/gene="BB0671"  
/note="similar to GP:1408273 percent identity: 98.76;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis operon protein (chex)"  
/protein\_id="AAC67022.1"  
/db\_xref="GI:2688605"  
/translation="MRIDYIEPLDAASSVLRMLLVENIMKPGKLSINOKIKGS  
VYGLAGSVG5I5IIDMDIT5ALFVASKLNEP5EDD5DET5EMAA5T5L5REV5N5IA  
GNFV5TL5HAK5F5D5IT5PA5F5IT5GEN5K5IS5K5EAL5V5LP5DGA5L5IEV5N5IA5IR5R  
V"  
5706..6146  
/gene="BB0672"  
5706..6146  
/gene="BB0672"

```

/note="similar to GP:1408274 percent identity: 97.95;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="chemotaxis response regulator (cheY-3)"
/protein_id="AACG7020.1"
/db_xref="GI:2688604"
/translation="MIKRTIIAADSSSKRGKINDTGTGPENVLIVDSVTFVKQIQTIT
FTSGSGIPIIDAAAGEAVATKYNHNPNDIVLTIDITMKMGDGLISINIEPKNKNA
VIMISALGKEIDVAKDCILKCAKPFIVKPIIDRAKVLQIVMSVFK"
/gene="BB0673"
/complement(6187. .6702)
/complement(6187. .6702)
/complement(6187. .6702)
/note="similar to GP:1408275 percent identity: 97.56;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AACG7020.1"
/db_xref="GI:2688603"
/translation="MKQIITIMLLALDPEPLNARLIDISTEKRADEIKKYSNLIILII
EKVEYINPSEIKENIKYKTEHPVSNIMKNTGNSNKKYKNTLIQILOSELIDKKK
ELAKIKIKININGIFPSHSLITKKGPFKELIYENNAEPKIFNLNITTFPLKNDKK
ISNKKIFPPR"
/gene="BB0674"
/complement(6699. .7745)
/complement(6699. .7745)
/gene="BB0674"
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0674"
/protein_id="AACG7027.1"
/db_xref="GI:2688610"
/translation="MKKSFIFLKATIIIFSELLIEELSLITFLPYIRPALIFLGF
LEDFITFLFLYKTKAYLSQRLEIVYRNNEFDDIHCLIPAFYSSVOLNIIIVAEH
TILPDMIMSLFELKFLFLNLDILEIYVNSKEKNLILIAAFYSSLSLIPTPTPI
ITSSKIVNSIPEKQENIENIKNSIENKAYIKKPKLIIKEDODIYFSQSEIVF
YVSSSEVRYEMKTEKQYIDKYORXSDSLIGLFTPLASPTPIAMFYKFPASFLT
NPILIMKLIQDDEPLEYKRIQIPTLSEKAYELANSFNMLLKEKLNKRSKIPLET
EKVKIINKDQEK"
/gene="BB0675"
/complement(7739. .8602)
/gene="BB0675"
/complement(7739. .8602)
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0675"
/protein_id="AACG7026.1"
/db_xref="GI:2688609"
/translation="MNTKPYFPLYLYLFNHSISLSAIEKEIILNLYKKNKTTA
TFINDESEIKELDIDVKKDKDITMTPEVLSGIEAIDENIYKPLFSEKELKNLDLII
PNFVKNSSAKLEVFYENINSIGVITPINKLFEGKNKSYIYLQKKNKALYSDDIIL
KNYKIKLILKLVKAYIKCEPKGIELTKNIESNSAISNDPFLDEKTAKLITSFPK
YETIOWMSPITSLTALFASAKRPKTKNNPVAGFICIDESWMSIKOSSREYDRIIK
ELSEIAKYNKK"
/gene="BB0676"
/complement(8599. .9261)
/gene="BB0676"
/complement(8599. .9261)
/note="similar to GP:142023 SP:94755 P1D:1004013
P1D:1222502 P1D:1204815 percent identity: 25.58;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphoglycolate phosphatase (gph)"
/protein_id="AACG7019.1"
/db_xref="GI:2688602"
/translation="MKACIKIPMDGTLVNSIMDIAFSNNSALSNLGSYKTELSKFNVA

```

Query Match	72.8%	Score 1210.6	DB 1	Length 13271
Best Local Similarity	98.5%	Prod. No. 2.6e-153		
Matches 1246	Conservative 0	Mismatches 9	Indels 10	Gaps 2
QY 401	taaatcttaccataaataataatataaaacgaataattataaagatt--tatata	458		
Db 1	TTAAATTTACTATATAAATAATTAATAAAGCAAAATTTTATAAAGATTATATATATA	60		
QY 459	aggagcttggttacatgaagaagaaagctaaagatttlatcttttlatatccact	518		
Db 61	AGGACTTGCGTTTACTAGTAAAGAAAGCAAGCAATATTTATTTTATATTCACCTG	120		
QY 519	ttctctttgtctcaagaagactgataagcagagggcttcaaaagggcagaagctgag	578		
Db 121	TTCTTTTGTGCTCAAGAGACTGATGATTAAGCAGAGAGGTTCTTAAAGCGCAGAGCTTGAG	180		
QY 579	aattagctcttgattttgcccagacttgcaagaagatccaagtcttaacttgacttacta	638		
Db 181	AATVAGTTTAAATTTTGCCGAGCTTGCAAGAGATCCAAAGTTCATGATCTTA	240		
QY 639	caaatatggttaattatgtatatacttcgggcgccttcggtatgtatgaagccggaataag	698		
Db 241	CAAAATTATGTTGATATATGATATTTGCGGCGCTTCGATATGTTAAGCGGGAAGATATAG	300		
QY 699	ttgtatgactcttggataaataatltgagagcgttttacttactctcttgcagaagttgcag	758		
Db 301	TTTGATGATCTTGCGATTAATAATTTTGAGCGGTTTACTTACTTCCTTGCAAGGTTGCAGG	360		
QY 739	cttaagcttaaaaaattcagtttgttcggcccgcgttgttaagagtgagtcgcaaaagtgacg	818		
Db 361	CTTAGCTTAAAAAATTCAGATGTTGGCGCCGCTGTTGTTAAGAGTGAAGTCAAAAAGATGACG	420		
QY 819	caaggtgatactatcttggggggtgaagaagtttgttcccaagcattcccaatcatctgta	878		
Db 421	CAGGATATCTATTTTATAGGGGTAAAGGTTTGTTCACAGCATTTCTCAATCATTCGTGA	480		
QY 879	tgatatabgccaacattaaattccctcttctatctcagaagggaaagtgaccaatctttag	938		
Db 481	TGATTTATGCGACCATTTTAAATTCCTTTTATTCAGGGGAAAGTGCATCAATTTTATAG	540		
QY 939	gcaaaagtccttatgataacatlaaaaccaagaaagaataaggatctgttttatagtt	998		
Db 541	GCAAAAGGCTTATATGATTAACATTTAAACCATGAAAGAAATTAAGGTATCTGTTTATAGTT	600		
QY 999	tagggatagagaatagactcttgaggtttatattggaagatgaaatgagcatggaatgctt	1058		
Db 601	TAGGATATGAGATACATCTTGAGGTTTATTTTGAAGATATCAATGCGATGCAATATGCTT	660		
QY 1059	attcctatgggtactttaaagtttaaagggctgagctgaattaaatttgcgtaactcc	1118		
Db 661	ATTCTATGGGTACTTAAAGTTTAAAGGTGCGGCTGATTTATTTTGGTCAATCTCTACT	720		
QY 1119	atacttccataataatcatccagaattatlaaagacgattgccaaattatcctcttgctt	1178		
Db 721	AATATTCCTAATTTATCAATCCAGAAATTTATTAAGAGAGATGTTCCAAATATATCTCTGCTT	780		
QY 1179	caagtaaaaaatgagatttaaggtctttagagtttcaaaagctcaacaagttcaaaagttaa	1238		
Db 781	CAGATTAATATGATTTTAAAGCTTTTAAAGTTTCAAAGTCAACACAGCTTCAAAAAGAGCAAA	840		
QY 1239	attcatattttatgtttaaagaatttaagaagttcttctatgataagagctlaagtglttcaatag	1298		
Db 841	ATTTCATCTTTTATGTTTAAAGATTTTAAAGATTTCTTTATGATTAAGTTGAGTGCTTTCATAG	900		
QY 1299	attctgatatatgacagtgagctctgtatlaaagtttatgagactagcggaactgataccc	1358		

Db 901 ATTCTGATATGACAGTGAAGCTGTAATTTAAAGTTATGAGACTAGCGAAGTAATCC 960  
 QY 1359 ttccgtaataataagcagccgaacttttaaaagagtttaagccttagaagaaaaattt 1418  
 Db 961 TTCGTAAATTAAGGCACGACGAACCTTTTAAAGCTTTTAAAGCTTAGAGAAAATTTT 1020  
 QY 1419 ccatcgctaaagcctcttccaaaacttctgtaaaaagcttagagtgaaaacccggaag 1478  
 Db 1021 CTAATGCTTAAGGCTCTTTCCAAACTTTGTAGAAAAGATTGAGATGAAAACCTGAGG 1080  
 QY 1479 aatcatctccgaaaaatagtgcttaaatatagtaagactaacctaaaggttgcctt 1538  
 Db 1081 AATCATCTCCGAAAAATAGTGTAAATTAATATGTAAGACATCAATAAAGCTTTCCTTT 1140  
 QY 1539 errorssmacatataataataggaataatagatagtgaaataatagatttgaaaag 1598  
 Db 1141 -----ACATATTTAAATTAATAGAAATAGATATATGAAATATTTGATTTGAAATG 1192  
 QY 1599 aagagcttttagagagtttttttgaagaagctcaaatcttctgagatagcttagaaga 1658  
 Db 1193 AAGAGCTTTTAGAGAGTTTATTTTGAAGAGCTCAAAATCTGTAGATATCTTGAAGAGA 1252  
 QY 1659 atact 1663  
 Db 1253 ATATT 1257  
 RESULT 3  
 AEO01167 10016 bp DNA BCT 15-DEC-1997  
 LOCUS Borrelia burgdorferi (section 53 of 70) of the complete genome.  
 DEFINITION AEO01167 AE000783  
 ACCESSION AEO01167.1 GI:268585  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Lyme disease spirochete.  
 Borrelia burgdorferi  
 Bacteria: Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
 burgdorferi group.  
 1 (bases 1 to 10016)  
 Fraser,C.M., Casjens,S., Huang,M.M., Sutton,G.G., Clayton,R.,  
 Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
 Gwinn,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D., Richardson,D.,  
 Peterson,J., Kerlavage,A.R., Quackenbush,J., Salzberg,S.,  
 Hanson,M., van Vugt,R., Palmer,N., Adams,M.D., Gocayne,J.,  
 Venter,J.C., et al.  
 Genomic sequence of a Lyme disease spirochaete, Borrelia  
 burgdorferi  
 Nature 390 (6660), 580-586 (1997)  
 JOURNAL 98065943  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 TITLE  
 Genomic sequence of a Lyme disease spirochaete, Borrelia  
 burgdorferi  
 JOURNAL 98065943  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 TITLE  
 Submitted (12-DEC-1997) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA  
 FEATURES  
 source  
 1..10016  
 /organism="Borrelia burgdorferi"  
 /db\_xref="taxon:139"  
 84..1217  
 /gene="BB0656"  
 84..1217  
 /gene="BB0656"  
 CDS  
 gene  
 /note="similar to SP:P54304 PID:1303805 GB:AL009126  
 percent identity: 25.62; identified by sequence  
 similarity: putative"  
 /codon\_start=1

gene  
 CDS  
 /transl\_table=11  
 /product="oxygen-independent coproporphyrinogen III  
 oxidase, putative"  
 /protein\_id="AAC67009.1"  
 /db\_xref="GI:2688591"  
 /translation="MRVLDPLPVELSLYINLSFCCNDFSIFNRIIEELKCHLILGHP  
 IIKTYIKHVDCLSRQONLKFETLSKXINLLEEEETLEIPYVPEFKLDE  
 RCITRIINLVOSFSLERKIVGIPETSYKKLILINIRKPPDINIDVTMMPLDK  
 SHLRDKLKLSTMPHEHICFSFICEBEFVLRDPDNSIDSEKLNFCALCESNGYI  
 NYETINFAELGHESRHNLKMWELKPHLGVLAVSLFCGDNNNYRALIRKTSFVK  
 ANNHVLTPELKLDEFFVYHFIGLGTIGVSLRALRFEFNEKQFQFIVNCSTLS  
 KKEFVDNIMLKGRERKLNFLYLVIIINHFDNFEKVLRLP"  
 complement(1204..1890)  
 /gene="BB0657"  
 complement(1204..1890)  
 /gene="BB0657"  
 /note="similar to GB:L77117 PID:1500500 percent identity:  
 39.53; identified by sequence similarity: putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ribose 5-phosphate isomerase (rpl)"  
 /protein\_id="AAC67008.1"  
 /db\_xref="GI:2688590"  
 /translation="MENOKILVAKHAIDHYIKSNMNLGIGTGTIYVIAIKYLSKIKS  
 GSKNLKFTTSSDPRKYLKSKQIPVSESKLNKLDIAIDGADLEILKSKLGM  
 GSAHLMKXIATNSETLITADETKYKLGTRKMPPIEVAQNAVGFITRLDENMLE  
 ATLRIKCEKKGPITIDNNNYILDVKNHVENPGETEFKRLPGLIEIGLFNHNKTRIV  
 YVQDKQIKKA"  
 2024..2785  
 /gene="BB0658"  
 2024..2785  
 /gene="BB0658"  
 /note="similar to SP:P31217 GB:U00096 PID:1651340  
 PID:1786970 percent identity: 59.76; identified by  
 sequence similarity: putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phosphoglycerate mutase (gpmA)"  
 /protein\_id="AAC67007.1"  
 /db\_xref="GI:2688589"  
 /translation="MLGDPMYKLVLRHGESEMNKENTFTGWTDLVLSKGDIEAVEA  
 GILLKQEGSPDIAPESLISLRANDITILIRELGOSIVYKTRMLNENHGYALGIN  
 KSETAKYGEKVILIRRSYDVPKMSLSDSDRRHKPDRIYKHIPRELPSRECLKDT  
 VARVIPWTDEIAKEVLEBKRVIAAHGNSLRALVTFPNLSBEDVLKLNIPGLIPLV  
 YELDKDNLPIKHYVLEDESKIRKMSVASQGLK"  
 complement(2853..4418)  
 /gene="BB0659"  
 complement(2853..4418)  
 /gene="BB0659"  
 complement(2853..4418)  
 /note="similar to GB:L77117 SP:Q57959 PID:1591243 percent  
 identity: 33.14; identified by sequence similarity:  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="lysyl-tRNA synthetase"  
 /protein\_id="AAC67006.1"  
 /db\_xref="GI:2688588"  
 /translation="MKTAMADFYAEKIRKGGPKNLYTVASGTPSGTVHIGNPREV  
 ISVDLVARALRDSGSKVRFTYSDNDVPRKRYKKNPEDELITTYRQAITRVDPORS  
 HKTSYARANIEFEKYLTPVGAIPNEDIFDIOFYSVAASQIRKALFELSEALNE  
 YRTSKLENNWYIPVCTGCNDTITVNNYNDHXYSEYCEGNOGLDITRTWARKL  
 PWRIDPMKRYEKVDPEPAKGDHSSGGSFDSKIVIKFQSGSPVTVQYDFISIKG  
 RGKGLSSSGDVIISLKLVLVETVETRLFAATKPNTEFSISFDLVIRKIEDYKFL  
 ERIVYGEVDVKEKRAKFAFRIYELSQVPMPSKRIPOVGFRIHLSVTSQAFENNKINIL  
 NYLKNVQEDQDKLINKINCAINMIDFAPEDKYSFLRSKFDMEILEENSKKAIHEL  
 LDPLKNEFVATEEDIQNEIYKISRNENIPALFFQIYKILIDREKGPKLAGFTIKII  
 GIDREKITSKYV"  
 complement(4505..5377)  
 /gene="BB0660"  
 complement(4505..5377)  
 /gene="BB0660"  
 complement(4505..5377)  
 /note="similar to GB:M14658 SP:P06616 PID:416295 PID:42767  
 PID:987641 percent identity: 37.85; identified by sequence

```

similarity: putative"
/codon_start=1
/transl_table=11
/product="GTP-binding protein (era)"
/protein_id="AAC67005.1"
/db_xref="gi:2688587"
/translation="MKSGEAAILGRPSNGKSTLNLSTGCHKISTISPIPOTTRANNIGK
1PTDRCGLITFDLPGLSKKKPKIAMKMHISIGVEYLITIDIOOKGSEEEK
MKEITKSKIKFVLYLNLKIDLNKTKIKITPTPLKGIEDSNITKISAEKINTEELK
NLIYFSEGLPYFQEYVTDQELNFRISITELREAILENLEELPYLYVIDPLELKE
KGSLEIRANIEFANESOKGIIVGNKGKIKSIGERARKTIAKIFETKCNLFQVKKKK
MNNKEDKILKRLN"
5467.5838
/gene="BB0661"
5467.5838
/gene="BB0661"
/annotation="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0661"
/protein_id="AAC67015.1"
/db_xref="gi:2688597"
/translation="MRNRFLAKVNFVEVCECGEKGVLGRSVCPNCAKIGDICPCNQC
RHVGVSVAFGECPCMHCKSPQELKEKFKRERARMYSDAAKTYRFLRFPINID
ILLYLFSSFLPYFLTYIVG"
5848.6258
/gene="BB0662"
5848.6258
/gene="BB0662"
/annotation="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0662"
/protein_id="AAC67014.1"
/db_xref="gi:2688596"
/translation="MGTIVFYLFSTFASFLGSSMDSKENVLKSTIFYDYVEVEPRP
YARKDTGFIANTHLKAVFNFDMKKMSYTFVFDKLLISOYALFIEVKKRFGATLV
TPLNIMLDGDSIIVLNKNIRILTKSYISYMK"
6278.6727
/gene="BB0663"
6278.6727
/gene="BB0663"
/annotation="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0663"
/protein_id="AAC67013.1"
/db_xref="gi:2688595"
/translation="MKKILKRFLLFVLSMPSLFADHPDYDEKIMNGVFKVJASNEB
1DRAGVGAAQVEYEGNGLMFLFKDDNLSPFMENTPLLEIAVSDSGIKELIYDLE
PYSRANVNSLYKVRPALDPRGSPFKQIKIGDKVHCFDYNALVE"
complement(6736..7419)
/gene="BB0664"
complement(6736..7419)
/gene="BB0664"
/annotation="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0664"
/protein_id="AAC67012.1"
/db_xref="gi:2688594"
/translation="WLHNTMLIARIMINTLTFYGMIIIFLLISCNHKNIOYDKRIKK
FLDKRKIEKIDSEDFIAFKDINNKEEYVIRSLNYSNKSRIEYFGVAKVDIN
TPKRIEISLSDSDYNNRVGSGWEIINNAERGINSLYIVKAEFAFDPLVAIDEB
IASLISIFKRIITTNENIDNNEENNNTSNESQPTLKQKNTSKESNNELKEDQIEE
EFLQPIKAO"
7462.8421
/gene="BB0665"
7462.8421
/gene="BB0665"

```

Query Match	26.2%	Score 435.8	DB 1	Length 10016
Best Local Similarity	97.6%	Pred. No. 8.9e-50		
Matches 456	Conservative 0	Mismatches 2	Indels 9	Gaps 1
QY 1	atgataatccttttcaaaaaaggcttttaatttcattccttagcaaggagcttgcta 60			
Db 9550	ATGATTAATCTTTTTCAAAAAGCTTTTATTTTCATTTCAGCAAGGATTTTGCTA 9609			
QY 61	attaaagatatttaagagatgacaacaaatttgaaagataattgaccttttaaaag 120			
Db 9610	ATTAAAGATATTTTAAGAGATGAACAAAATTTGAAAGATAATTGATCTTTAGTAAAG 9669			
QY 121	attcttccttagaagcaatgagggtcttccttccttgagcttttaagtgatccaagat 180			
Db 9670	ATTTCCTTTTAGGAGCATGACAGGTTTCTTTGGGTTTATTAAGTCATTCAAGAT 9729			
QY 181	tttataatcctttttaaagaatgagtttattatgataatccttccaagagaatttaty 240			
Db 9730	TTTATATCTTTTAAAGAAATGAGATTTATGTAAATCTTCAAGACATTTTATG 9789			
QY 241	attcctttaaataatgagatata-----taatgaatcctttgagtgtaaggtcaatc 291			
Db 9790	ATTCCTTTAAATGAAGCGTATTATATGATGATCAATCAATCTTTTGATGTTAAGTCATC 9849			
QY 292	tttttgcatgctttttaaataaacaatgagcctttaaactccctggtaaagataaaga 351			
Db 9850	TTTTGCTATGCTTTTATTAATAAACATGCGCTTTAACTACTGCTGTTAAAAAGA 9909			
QY 352	ttatataatccttgtaagagtgatctttaaagagcaagtgtaataatttact 411			
Db 9910	TTGTTATCTCTGTTGAACGCTGTATCTTAAAGAGCAAGTGAATAATTAAATTTACT 9969			
QY 412	aataaaataattaaagaagaatttataaagaattatata 458			
Db 9970	AATAAAATATATTAAAAACGAAAATTTTATAAAAAGATTATATATA 10016			

RESULT 4

BBCHEAGEN 332 bp DNA BCT 20-JAN-1998

LOCUS B.burgdorferi chea gene.

DEFINITION X91907

ACCESSION X91907.1 GI:1019753

VERSION

KEYWORDS chea gene; chea protein; chemotactic response protein; histidine kinase.

SOURCE Lyme disease spirochete.

ORGANISM Borrelia burgdorferi

REFERENCE 1 (bases 1 to 332)

AUTHORS Truebe,G.A., Old,I.G., Saint Giron,S.I. and Johnson,R.C.

TITLE A chea cheW operon in Borrelia burgdorferi, the agent of Lyme disease

JOURNAL Res. Microbiol. 148 (3), 191-200 (1997)

MEDLINE 96438936

REFERENCE 2 (bases 1 to 332)

AUTHORS Old,I.G.

TITLE Direct Submission

JOURNAL Submitted (03-OCT-1995) I.G. Old, Institut Pasteur, Bacteriology & Mycology, 28 rue du Dr Roux, 75724 Paris cedex 15, FRANCE

FEATURES

source

1..332

```

/organism="Borrelia burgdorferi"
/strain="212"
/db_xref="taxon:139"
/clone_lib="Cos1"
1..202
/gene="orf"
<1..202
/gene="orf"
/codon_start=2
/transl_table=11
/protein_id="CAA63001.1"
/db_xref="GI:1019754"
/db_xref="SPTREMBL:044876"
/translation="DSDDSESYFVKYETSGTESLRKLAHETFKRVYLKREKISTAE
GSFQNFVEKIESEKPESSPKN"
repeat_region
205..256
/feature="hairpin"
RBS
258..261
/gene="chea"
270..332
/gene="chea"
270..332
/gene="chea"
/feature="chemotactic response"
/codon_start=1
/transl_table=11
/product="histidine kinase"
/protein_id="CAA63002.1"
/db_xref="GI:1019755"
/db_xref="SWISS-PROT:044737"
/translation="MEIDLLENBELLGVFEFAQN"
BASE COUNT      130 a      37 c      61 g     104 t
ORIGIN

```

```

Query Match      18.9%; Score 314; DB 1; Length 332;
Best Local Similarity 97.6%; Pred. No. 2.5e-33;
Matches 332; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1297 agattctgatatgacagtgagtcgtatttaaaagttatgagaactagcggaactgaatc 1356
|||||
DB 1 AGATTCTGATATTGACAGTAGCTGTGATTTAAAGTTATGACGACTGACGGAACGTGAATC 60
|||||

QY 1357 ccttcgtaaatlaaagcagacgaacttttaaaagtttaaaagcttagagagaaaaaat 1416
|||||
DB 61 CCTTCGTAATTTAAAGCAGACGAAACCTTTTAAAGAGCTTTTAAAGCTTAGAGAAAAAAT 120
|||||

QY 1417 ttctatcgctgaaggtcttccaaaacttctglaaagaagatgagagtgaaaaactga 1476
|||||
DB 121 TTCTATCGCTGAAGGCTCTTCCAAAACCTTTGTAGAAAAGATTGAGAGTCAAAAACCTGA 180
|||||

QY 1477 agaatacttcggaataaataagtttaaatataatgtaagctactactaaaggttgc 1536
|||||
DB 181 AGAATCATCTCCGAAAATAAGTTAAATTAATATGCTAAAGCTTAAAGGTTTGC 240
|||||

QY 1537 tteroyssnacatatataataggaatagatcataggaataattagatttgaaaa 1596
|||||
DB 241 TT-----ACATATTAAATAATAGGAATAGTATATGGAATATTAGATTGGAAAA 292
|||||

QY 1597 tgaagagctttagaggttttttttgaagaagctcaaat 1636
|||||
DB 293 TGAAGAGCTTTTAGAGGTTTGTGGAAGAACTCAAAAT 332
|||||

```

```

RESULT 5
LOCUS      BBU28962      2700 bp      DNA
DEFINITION Borrelia burgdorferi histidine kinase (chea) gene, complete cds.
ACCESSION U28962
VERSION   U28962.1 GI:1113814
KEYWORDS
SOURCE
ORGANISM  Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia

```

```

REFERENCE      1 (bases 1 to 2700)
AUTHORS       Trueba,G.A., Old,I.G., Saint Girons,I. and Johnson,R.C.
TITLE         A chea chew operon in Borrelia burgdorferi, the agent of Lyme
              disease
JOURNAL       Res. Microbiol. 148 (3), 191-200 (1997)
MEDLINE       98438936
REFERENCE      2 (bases 1 to 2700)
AUTHORS       Trueba,G.A. and Johnson,R.C.
TITLE         Direct Submission
JOURNAL       Submitted (12-JUN-1995) Gabriel A. Trueba, Microbiology, University
              of Minnesota, 420 Delaware Street S.E., Minneapolis, MN 55455, USA
              On Dec 12, 1995 this sequence version replaced gi:984552.
FEATURES
  source
    1..2700
    /organism="Borrelia burgdorferi"
    /strain="CT-1"
    /db_xref="taxon:139"
    95..2689
    /gene="chea"
    95..2689
    /gene="chea"
    /codon_start=1
    /transl_table=11
    /function="involved in chemotaxis"
    /product="histidine kinase"
    /protein_id="AB96835.1"
    /db_xref="GI:1113815"

```

```

gene
CDS

```

```

BASE COUNT      988 a      278 c      507 g      927 t
ORIGIN

```

```

Query Match      9.9%; Score 164.4; DB 1; Length 2700;
Best Local Similarity 95.3%; Pred. No. 1.9e-13;
Matches 183; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 1472 cctgaagaatcatctccgaaaaatlaagtttaaatatgtaaaagctactaaaggt 1531
|||||
DB 1 CCTGAAGAATCATCTCCGAAAATAAGTTAAATTAATATGTAAGCTACTTAAAGGT 60
|||||

QY 1532 ttgcttctteroyssnacatatataataggaataagatataaggaataattgattg 1591
|||||
DB 61 TTGCTT-----ACATATTAAATAATAGGAATAGTATATGGAATATTAGATTGG 112
|||||

QY 1592 gaaatgaagagctttagaggttttttttgaagaagctcaaatctgtagatattc 1651
|||||
DB 113 GAAATGAGAGATTTTAGAGGTTTTTTTGAAGAACTCAAAATCTTGTGATATCTCT 172
|||||

QY 1652 gaagaagaatatt 1663
|||||
DB 173 GAAGAGAAATATT 184
|||||

```

```

RESULT 6
LOCUS      A24079      1140 bp      mRNA
DEFINITION T. hyodysenteriae 43 kd sheath protein gene.

```

ACCESSION A24079  
 VERSION A24079.1 GI:904392  
 KEYWORDS  
 SOURCE Brachyspira hyodysenteriae.  
 ORGANISM Brachyspira hyodysenteriae; Bacteriia; Spirochaetales; Brachyspiraceae; Brachyspira.  
 REFERENCE 1 (bases 1 to 1140)  
 AUTHORS Koopman, M. B. H. and Kusters, J. G.  
 TITLE Treponema hyodysenteriae vaccine  
 JOURNAL Patent: EP 0534526-A 1 31-MAR-1993;  
 DUPHAR INTERNATIONAL RESEARCH B.V.  
 FEATURES  
 source location/qualifiers  
 1..1140  
 /organism="Brachyspira hyodysenteriae"  
 /strain="C5"  
 /db\_xref="taxon:159"  
 /cell\_line="CBS 513.91"  
 -35\_signal 27..32  
 /note="putative"  
 -10\_signal 50..55  
 /note="putative"  
 RBS 86..91  
 /note="putative"  
 sig\_peptide 103..159  
 CDS 103..1065  
 /codon\_start=1  
 /transl\_table=11  
 /product="endoflagellar sheath protein"  
 /protein\_id="CAA01715.1"  
 /db\_xref="GI:904393"  
 /db\_xref="SWISS-PROT:P32520"  
 /translation="MKKLFVLTSLFIAASAYGLNSTLIDFALTGNDNLQAGEGDT  
 NEVVPVAENLVNDNMVWLNESARLTENRRNSYVTVNDSKNGAMEGKYLGVVHF  
 PLAAANSYALVAPYELFMGAGDGTKEGKYIHNGEIKSISWYGGNYLISYF  
 VNLQERBELKSYPMGYTFNGMROVRENREYLPNVRDRVLVREPLYPRMIPSKLD  
 SLGFYRKDTKGDFITVVKDVTLEYDVVVDVEEDIDDEATWDLTKTENDKQATIES  
 ARIRQAEIRDLQRIGDRIADGGAANTGAADTGAOEOAO"

BASE COUNT 390 a 152 c 233 g 365 t  
 ORIGIN

Query Match 6.5%; Score 108.6; DB 6; Length 1140;  
 Best Local Similarity 49.4%; Pred. No. 6e-06;  
 Matches 369; Conservative 0; Mismatches 369; Indels 9; Gaps 3;  
 QY 581 ttatgtttagattttgctgagagcttgcagaagatccaaagttcaactgaacttgaattcttaca 640  
 DB 136 tttatcgctgcactgcttaccgttttaacaaactcaacttttgattgatttgcgttttaaca 195  
 QY 641 aattatgttgaattatgtatattcggcgcttctgtgattgttaagcggagaagatcgtt 700  
 DB 196 ggtaattgcgtatatttactaacgttgagaaagtgatcaaatgaattgtccagttgca 255  
 QY 701 gtatagcttgggataaataattggagcgttttactactcttctgcaagttgcagact 760  
 DB 256 gaaaaacttttaataatgatgagtgagtgatgattgaatgaattcctgatttaacagag 315  
 QY 761 taagttaaaattcagttgttgcgcgcgtgtgttaagagtgagtcacaaaagttcgca 820  
 DB 316 aatcgcgaaaattcttattgttactaacgtagacagtaaaagtgtaaacacgttgcgttgga 375  
 QY 821 ggtagaacattttgggggtaagagttgttttcaaacgattatcctc---aatacttgt 877  
 DB 376 gcagagtaaggttcttgcttgaagagtagactttccctttagcagcttggaaacgtttgct 435  
 QY 878 atgattatgccacacattaaatctcctttatttcaaggagaaagtggcaatcaattc--- 934  
 DB 436 ttatgtaaaacagatattgaacttgaatgctgagcggtgctgactgatttaataataga 495  
 QY 935 ttagagaaaggtctatttataacattaaacatgaagaagaattgaattatcgtttat 994  
 DB 496 gaaggtaaagggttattatcattacattgacgttggcgaaatcaattcttattggttgggtttat 555

QY 995 agttaaggatgatgatgatccttgaagtttatttcaagatataatgcatggaatat 1054  
 DB 556 gcagctaatctatttatttatttgcgaacttacaataatggaatttgcgaataat 615  
 QY 1055 gcttttctatggtactttaaagtttaagggttgccgatttaatttgcgaatcct 1114  
 DB 616 tctttatccaatggctgcttttacttcaacgggttgagacaaatgattggaaacaca 675  
 QY 1115 aactatattccaatatataatccaggaattataaagacagatgttccaaatcctct 1174  
 DB 676 gattacttaccatattgttgcgcacagatattagtaagaaacgtct---tatcttga 732  
 QY 1175 gcttcaagtaaaatgagatttaaggctttagagtttcaagtcacacagttcaagtt 1234  
 DB 733 atgattcccttctgttaatttatttatttctttatgagtaaaagacttaaaagcc 792  
 QY 1235 aaaaattcatctttatttatttgaagtttaagtttctttagataagcattgttca 1294  
 DB 793 ggagatttcattcattcattcattcattcattcattcattcattcattcattcattcatt 852  
 QY 1295 atagattcgtatgtacagagtgact 1321  
 DB 853 ttggaagaagatattcagcattgact 879

RESULT 7  
 THFLA1 1260 bp DNA BCT 01-DEC-1993

LOCUS T.hydroysenteriae fla1 gene.  
 DEFINITION X63006 S38343  
 ACCESSION X63006.1 GI:433523  
 VERSION  
 KEYWORDS fla1 gene; fla1 protein; flagellar sheath; flagellin; periplasmic protein.

SOURCE Brachyspira hyodysenteriae.  
 ORGANISM Brachyspira hyodysenteriae; Bacteriia; Spirochaetales; Brachyspiraceae; Brachyspira.  
 REFERENCE 1 (bases 1 to 1260)  
 AUTHORS Kusters, J. G.  
 TITLE Direct Submision  
 JOURNAL Submitted (04-NOV-1991) J. G. Kusters, Inst. of Infectious Diseases and Immun., P.O.Box 80.165, 3508 TD Utrecht, THE NETHERLANDS

REFERENCE 2 (bases 1 to 1260)  
 AUTHORS Koopman, M. B., de Leeuw, O. S., van der Zeijst, B. M. and Kusters, J. G.  
 TITLE Cloning and DNA sequence analysis of a Serpulina (Treponema) hyodysenteriae gene encoding a periplasmic flagellar sheath protein  
 JOURNAL Infect. Immun. 60 (7), 2920-2925 (1992)

MEDLINE 92307926  
 FEATURES  
 source location/qualifiers  
 1..1260  
 /organism="Brachyspira hyodysenteriae"  
 /strain="C5"  
 /db\_xref="taxon:159"

-35\_signal 27..32  
 -10\_signal 50..55  
 RBS 86..91  
 gene 103..1065  
 /gene="fla1"  
 sig\_peptide 103..159  
 CDS 103..1065  
 /gene="fla1"  
 /gene="fla1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="flagellin Fla1"  
 /protein\_id="CAA4735.1"  
 /db\_xref="GI:433524"

/db\_xref="SWISS-PROT:P32520"  
 /translation="MKKLFVLTSLFIAASAYGLNSTLIDFALTGNDNLQAGEGDT  
 NEVVPVAENLVNDNMVWLNESARLTENRRNSYVTVNDSKNGAMEGKYLGVVHF  
 PLAAANSYALVAPYELFMGAGDGTKEGKYIHNGEIKSISWYGGNYLISYF  
 VNLQERBELKSYPMGYTFNGMROVRENREYLPNVRDRVLVREPLYPRMIPSKLD  
 SLGFYRKDTKGDFITVVKDVTLEYDVVVDVEEDIDDEATWDLTKTENDKQATIES  
 ARIRQAEIRDLQRIGDRIADGGAANTGAADTGAOEOAO"

```

mat_peptide      160. .1062
                  /gene="f1a1"
repeat_region    1004. .1030
                  /product="f1a1"
repeat_region    1157. .1175
                  /rpt_type=INVERTED
BASE COUNT      437 a 158 c 245 g 420 t
ORIGIN
Query Match      6.5%; Score 108.6; DB 1: Length 1260;
Best Local Similarity 49.4%; Pred. No 5.9e-06;
Matches 369; Conservative 0; Mismatches 369; Indels 9; Gaps 3;

QY 581 ttatgcttagatttgcgcgagcttgcaagatccaaatcactagactgattcttaca 640
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 136 ttatgcttagatttgcgcgagcttgcaagatccaaatcactagactgattcttaca 195

QY 641 aatagcttagatttgcgcgagcttgcaagatccaaatcactagactgattcttaca 700
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 196 gtaatgcttagatttgcgcgagcttgcaagatccaaatcactagactgattcttaca 255

QY 701 gtaatgcttagatttgcgcgagcttgcaagatccaaatcactagactgattcttaca 760
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 256 gtaatgcttagatttgcgcgagcttgcaagatccaaatcactagactgattcttaca 315

QY 761 taccgttaaaatcagcttgctgcgcgagcttgcaagatccaaatcactagactgattcttaca 820
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 316 taccgttaaaatcagcttgctgcgcgagcttgcaagatccaaatcactagactgattcttaca 375

QY 821 gtagatacttcttgaggagcttgctgcgcgagcttgcaagatccaaatcactagactgattcttaca 877
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 376 gtagatacttcttgaggagcttgctgcgcgagcttgcaagatccaaatcactagactgattcttaca 435

QY 878 atgattatgcacacatcctaataatccttcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 436 atgattatgcacacatcctaataatccttcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 495

QY 935 ttggcgaagcttcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 496 ttggcgaagcttcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 555

QY 995 agcttaagagcttcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 556 agcttaagagcttcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 615

QY 1055 gcttattcttcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 616 gcttattcttcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 675

QY 1115 aactatacttcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 1174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 676 aactatacttcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 732

QY 1175 gcttcaagcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 1234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 733 gcttcaagcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 792

QY 1235 aaaaattcttcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 1294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 793 aaaaattcttcttcaatcgaagagcttgcaagatccaaatcactagactgattcttaca 852

QY 1295 atagattctgatacttgcaagagcttgcaagatccaaatcactagactgattcttaca 1354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 853 atagattctgatacttgcaagagcttgcaagatccaaatcactagactgattcttaca 912

RESULT 8
PEMALIP3/c 67970 bp DNA INV 15-DEC-1999
LOCUS PFMALIP3 67970 bp DNA INV 15-DEC-1999
DEFINITION Plasmodium falciparum MALIP3, complete sequence.
ACCESSION AL031746

```

```

VERSION          AL031746.9  GI:6594243
KEYWORDS         HTG.
SOURCE           malarial parasite P. falciparum.
ORGANISM         Plasmodium falciparum
REFERENCE        Eukaryota: Alveolata: Apicomplexa: Haemosporida: Plasmodium.
AUTHORS          Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
                  and Barrett,B.
TITLE            Direct Submission
JOURNAL          Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium,
                  The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
                  CB10 1SA, UK
COMMENT          On Dec 16, 1999 this sequence version replaced gi:5763807.
                  For more information about this sequence or the malaria project,
                  see http://www.sanger.ac.uk/projects/p.falciparum. IMPORTANT: This
                  sequence is unfinished and does not necessarily represent the
                  correct sequence. Work on the sequence is in progress and the
                  release of this data is based on the understanding that the
                  sequence may change as work continues. The sequence may be
                  contaminated with foreign sequence from E.coli, yeast, vector,
                  phage etc.

FEATURES         Location/Qualifiers
source           1. 67970
                  /organism="Plasmodium falciparum"
                  /strain="3D7"
                  /db_xref="taxon:5833"
                  /chromosome="1"
                  /complement(join(1748. 2598,2748. 2848,2990. 3276))
                  /gene="MALIP3.01"
                  /complement(join(1748. 2598,2748. 2848,2990. 3276))
                  /gene="MALIP3.01"
                  /note="MALIP3.01, conserved hypothetical protein, len: 412
                  aa, similarity: UPF0006 family, eg to
                  YB055C/YB0512/YB0511, YB53, YB57 (418 aa), fasta
                  scores: Opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa
                  overlap)"
                  /codon_start=1
                  /product="conserved hypothetical protein, UPF0006 family"
                  /protein_id="CAB63556.1"
                  /db_xref="GI:6594244"
                  /translation="MKLVFHYIYINLVFYSIIIFLKSNSLKYINDLRVTSVNVKKV
                  LQIKRSLNLRNHNIRKMDNESFIDIGSLNDKMDGYNSKHNKNDONLVNRK
                  NNNWDKIITCTCLAEIDKSLKICETDPEKFLYSAGVHPNCKYEFIDKNEEKE
                  ITAKKEEPIKRYKFNQOVNSKMNENKRCICGEGKMNINLEIKNDLTPGFY
                  NEKDKEITLKKKIKIYPRIVICIGSIDPRLRCSYTIQIKYITPDKLVOMFN
                  LPMFLHMRNCSSEFPKIVDILKFLPEKNGVIVISFDKEDIVHIVQVYNLVIGVNG
                  CSLKSLEINAVKRIPLNLLETPDAPWCGVKRTKTHASVEYIKDYERAVYNLKKIKN
                  IIRCDNITIFKERNEPYNA"
                  /complement(2599. 2610)
                  /note="potential splice acceptor sequence"
                  /complement(2742. 2747)
                  /note="potential splice donor sequence, atg/gtataa"
                  /complement(2849. 2861)
                  /note="potential splice acceptor sequence"
                  /complement(2984. 2989)
                  /note="potential splice donor sequence, aaa/gtataaa"
                  /gene="MALIP3.02"
                  /complement(5005. 5496)
                  /note="MALIP3.02, hypothetical protein, len: 163 aa,
                  contains possible signal sequence"
                  /product="hypothetical protein, MALIP3.02"
                  /codon_start=1
                  /protein_id="CAB63557.1"
                  /db_xref="GI:6594245"
                  /translation="MKLVNRFVVICPIIIFPFIINSVYVGNRRNNRNINHEFNAX
                  AMRKLISGEINSLKINDGDELKIKLNDKEDKSTKDKSISFISNEEYISGTDLFR
                  KQEIENAVKRIIEDROEFYILNNDETENIATRFVLENNDELIVSFKOSLIDIIOS
                  LNN"
                  /note="possible cent, region of very high [A+T] content"
                  /complement(8020. 14884. 20352)

```







```

Db 8166 TTTATTATATATATATATTTGTTAATTTATATATTTATATATTTTATATAT 8107
      |||||
Oy 301 tgccttaataaacaatgcgttaactatccctgtaagaataaaagattattc 360
      |||||
Db 8106 TTTATTATTTATTTATTTCCATTATATATATATATATATATATATATATAT 8049
      |||||
Oy 361 tcttgaagggtctatctaaaggagcaagctgataaatttacttaataaat 420
      |||||
Db 8048 TTTCTTATATTTATTTATTTATATATATATATATATATATATATATATAT 7989
      |||||
Oy 421 aattaaaaaagaaaaattatataaagattatataaaggagctgttacaagaa 480
      |||||
Db 7988 ATATATATTTATATATATATATATATATATATATATATATATATATATAT 7929
      |||||
Oy 481 gaaagctaaagctattatttttttattatccactgtt 520
      |||||
Db 7928 ATTAATATATATATATATATTTATTTATATATATATATATATATATATAT 7889
      |||||

RESULT 9
LOCUS AC006986 173957 bp DNA PRI 14-OCT-2000
DEFINITION Homo sapiens BAC clone RP11-155J5 from Y, complete sequence.
ACCESSION AC006986
VERSION AC006986.3 GI:10801454
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 173957)
AUTHORS Sulston,J.E. and Waterston,R.
JOURNAL Toward a complete human genome sequence
MEDLINE Genome Res. 8 (11), 1097-1108 (1998)
99063792
REFERENCE 2 (bases 1 to 173957)
AUTHORS Joshi,C., Mohlman,P. and Phillips,A.
JOURNAL The sequence of Homo sapiens BAC clone RP11-155J5
TITLE Unpublished
REFERENCE 3 (bases 1 to 173957)
AUTHORS Waterston,R.H.
JOURNAL Direct Submision
TITLE Submitted (05-MAR-1999) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 173957)
AUTHORS Waterston,R.H.
JOURNAL Direct Submision
TITLE Submitted (05-MAY-1999) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 173957)
AUTHORS Waterston,R.
JOURNAL Direct Submision
TITLE Submitted (22-OCT-1999) Department of Genetics, Washington
UNIVERSITY, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 173957)
AUTHORS Waterston,R.
JOURNAL Direct Submision
TITLE Submitted (21-DEC-1999) Department of Genetics, Washington
UNIVERSITY, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 7 (bases 1 to 173957)
AUTHORS Waterston,R.
JOURNAL Direct Submision
TITLE Submitted (14-OCT-2000) Department of Genetics, Washington
UNIVERSITY, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 14, 2000 this sequence version replaced gi:4753246.
COMMENT -----Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu

```

----- Summary Statistics  
Center project name: H\_NH0155035

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skalecky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Fengen,E., Tatenio,M., Catanesi,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)  
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the right is RP11-160K17. Actual start of this clone is at base position 1 of RP11-155J5; actual end is at base position 173957 of RP11-155J5.

## FEATURES

```

source
Location/Qualifiers
1..173957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="Y"
/map="y"
/clone="RP11-155J5"
/clone_lib="RPCI-11"
/rpt_family="AT_r1ch"
1281..1435
/rpt_family="Alu"
1438..1582
/rpt_family="(TA)n"
1621..2154
/rpt_family="(TA)n"
2160..7755
/rpt_family="L1"
7758..10708
/rpt_family="L1"
11258..11377
/rpt_family="L2"
11934..12039
/rpt_family="(TTATATG)n"
repeat_region
12046..12950
/rpt_family="L1"
repeat_region
15316..15479
/rpt_family="(TA)n"
23465..23506
/rpt_family="AT_r1ch"
repeat_region
23633..23653
/rpt_family="AT_r1ch"
23961..23996
/rpt_family="(TG)n"

```

repeat_region	24034. .24234	/rpt_family="Alu"
repeat_region	27018. .27131	/rpt_family="Alu"
repeat_region	27132. .27154	/rpt_family="Alu"
repeat_region	27155. .27312	/rpt_family="(CA)n"
repeat_region	27324. .27617	/rpt_family="Alu"
repeat_region	29398. .29681	/rpt_family="Alu"
repeat_region	29682. .29720	/rpt_family="Alu"
repeat_region	29721. .29736	/rpt_family="(CA)n"
repeat_region	35366. .35796	/rpt_family="Alu"
repeat_region	35775. .36379	/rpt_family="Retroviral"
repeat_region	36749. .36868	/rpt_family="Retroviral"
repeat_region	37699. .37719	/rpt_family="AT-rich"
repeat_region	38833. .38860	/rpt_family="AT-rich"
repeat_region	39860. .40054	/rpt_family="MIR"
repeat_region	40141. .40187	/rpt_family="(CA)n"
repeat_region	40832. .41426	/rpt_family="L2"
repeat_region	41431. .41523	/rpt_family="Alu"
repeat_region	41555. .41587	/rpt_family="Alu"
repeat_region	41589. .44483	/rpt_family="r-rich"
repeat_region	43510. .43651	/rpt_family="Alu"
repeat_region	43652. .43952	/rpt_family="L2"
repeat_region	44279. .44197	/rpt_family="L1"
repeat_region	49198. .49332	/rpt_family="Alu"
repeat_region	49333. .50487	/rpt_family="Alu"
repeat_region	51792. .52073	/rpt_family="L1"
repeat_region	52074. .52105	/rpt_family="Alu"
repeat_region	52192. .52482	/rpt_family="(CA)n"
repeat_region	52511. .52829	/rpt_family="MIR1-type"
repeat_region	53366. .53426	/rpt_family="MIR4-group?"
repeat_region	54804. .54834	/rpt_family="L2"
repeat_region	55195. .55566	/rpt_family="AT-rich"
repeat_region	55792. .55940	/rpt_family="L2"
repeat_region	56086. .56299	/rpt_family="L2"
STS	56680. .56604	/rpt_family="L2"
repeat_region	57105. .57136	/rpt_family="L2"
repeat_region	57159. .57244	/rpt_family="(TG)n"

```
repeat_region      /rpt_family="(TA)n"
                    57276.57309
                    /rpt_family="(TG)n"
```

Query Match	5.4%;	Score 89.4;	DB 9;	Length 173957;
Best Local Similarity	49.9%;	Pred. No. 0.0012;		
Matches 225;	Conservative 0;	Mismatches 226;	Indels 0;	Gaps 0;

[illegible]

RESULT 10  
PFMAL4P1.0  
MPCOMMENT  
Sequence split into 4 fragments LOCUS PFMAL4P1 Accession AL034557

Sequence split into 4 fragments	LOCUS PFMAL4P1	Accession AL034557
Fragment Name	Begin	End
PFMAL4P1_0	1	110000
PFMAL4P1_1	100001	210000
PFMAL4P1_2	200001	310000
PFMAL4P1_3	300001	392633
LOCUS PFMAL4P1	392633 bp	DNA
DEFINITION	plasmidium falciparum chromosome 4	HTG 11-AUG-1999
PROGRESS ***	in unordered pieces.	*** SEQUENCING IN

ACCESSION AL034557  
VERSION AL034557.7 GI:5731897  
KEYWORDS HTG; HTGS\_PHASE1.

KEYWORDS HIG; HIGS\_PHASE1.  
SOURCE malaria parasite P. falciparum.

ORGANISM *Plasmodium falciparum*

Eukaryota; Alveolata; Apicomplexa

REFERENCE 1 (bases 1 to 392633)

**AUTHORS** Bowman, S., Churcher, C., Harris,

and Barrell, B.

TITLE	Direct Submission
TOWNSHIP	Submittal #3 (17-DEC-1009) D 43104

JOURNAL  
Submitted (17-DEC-1998) P. falci  
the Center for Wolfram Tru

The Sanger Centre, Wellcome Trust  
CR10 1SA UK

COMMENT CB10 ISA, UK  
On Aug 12 1999 this sequence

**COMMENT** On Aug 12, 1999 this sequence was  
 For more information about this

For more information about this  
see <http://www.sanger.ac.uk/Proc>

see <http://www.sanger.ac.uk/Pro>  
sequence is unfinished and does

sequence is unfinished and does not contain a correct sequence. Work on the sequence.

correct sequence. Work on the release of this data is based on

release of this data is based on the understanding that the sequence may change as work continues.

sequence may change as work con

On Aug 12 1999 this sequence version replaced g1:5531346. For more information about this sequence or the Malaria Project, see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum). IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be

and Barrell, B.  
Direct Submission  
Submitted (17-DEC-1998) P. falciparum Genome Sequencing Consortium,  
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA, UK

malaria parasite P. falciparum.  
Plasmodium falciparum  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
1 (bases 1 to 392633)  
Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M.

plasmodium falci parum chromosome 4 strain 3D7, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in unordered pieces.  
AL034557  
AL034557.7 GI:5731897  
HTG: HTGS\_PHASE1.

Primer	Sequence	Position	Size (bp)	HTG
P1_0	1	110000		
P1_1	100001	210000		
P1_2	200001	310000		
P1_3	300001	392633		
PEM14P1	392633	DNA		
				11-AUG-1999

split into 4 fragments	LOCUS FFMAL4P1	Accession AL034557
at Name	Begin	End
1	1	1000
2	1000	2000
3	2000	3000
4	3000	4000

gaaagctaaagaattatattttttaa 510  
| | | | |  
ATTATGAAAATAAATAATTTCATTA 1673

aatlaaaaaacgaaattttatataaagattatataaagggttggttacatgaaa 479  
 -  
 ATATGAAACCAATATATATTTTCATGATTCATATATGAAAAATTAATTAATTCATGATTT 1704

TTTATATTATGAGAAATAAATATATATTATATATATTTATATATGAAAATTAATATATTTTATA 1824  
TTTGtgaaggggtatcttaaggagcaagttgataaltaactlactataaaaa 419  
TTGAATAATAAATATTTTCCTGATTCATCATATGAAAATTAATATATTTTCATGTATFCA 1764

TTTTATTCGTTAAATAAATATTTATTTATTTATATGAAATAAATATTTTATAT 1884  
 TTTTCTTAAATAAACATCGCTTAACATCTGTAAGATAAAGATATTTAT 359

[illegible]

atttcttttaggaagcaatgagggtlcttccttggggtttcatagaatcaga 179  
- - - - -  
TATATTATATATTTATATATAAAAACAAATATATAATATATTATATATAAAACAAAT 2004

```

atttaagatatatttaagagatgaacaaatttgaagataatttagactctttagtaaa 119
222; Conservative 0; Mismatches 226; Indels 0; Gaps 0;
|||||
ATAAATATATTTTATATTTTAAATAAAAAATAATATATTTTATTTTGGTAAAAATTA 2064

```

```

Region      57276      .57309
            /rpl_family="rpln"
5.4% ; Score 89.4 ; DB 9 ; Length 173957 ;
Similarity 49.9% ; Pred. No. 0.0012 ;

```

```
/rot familyv="(TA)n"
```

Page 12



AC079314 242513 bp DNA HTG 09-AUG-2001  
 LOCUS Homo sapiens chromosome Xp clone RP11-386120, WORKING DRAFT  
 DEFINITION  
 AC079314 25 GI:15042762  
 AC079314.25  
 SEQUENCE, 28 unordered pieces.  
 AC079314.25  
 VERSION  
 AC079314.25  
 KEYWORDS  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE  
 1 (bases 1 to 242513)  
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
 Alsprouks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbara, J.,  
 Benton, J., Blinze, K., Blankenburg, K., Bonnin, D., Bouch, J.,  
 Bowe, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C.,  
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,  
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
 Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
 Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
 Davy, Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
 Dem, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
 Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
 Elhaj, C., Escotto, M., Falis, T., Ferraguto, D., Flaeg, N., Ford, J.,  
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
 Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S.,  
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
 Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C.,  
 Hollins, B., Homs, F., Howard, S., Huber, J., Hulik, S., Hume, J.,  
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
 Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
 Kovar, C., Kratovic, J., Kuresh, A., Landry, N., Leal, B., Lewis, L.C.,  
 Lewis, L., Li, J., Li, Z., Lichte, O., Liu, C., Liu, J., Liu, W.,  
 Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,  
 Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,  
 Ogutu, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojudo, I., Rolfe, M.,  
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshari, N.,  
 Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,  
 Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K.,  
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R.,  
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,  
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
 Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,  
 Worley, K., Wu, C., Wu, Y., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 242513)  
 Worley, K.C.  
 Direct Submission  
 Submitted (27-AUG-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 31, 2001 this sequence version replaced gi:1491136.  
 COMMENT  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc.help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: HGBA  
 Center clone name: RP11-386120  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; M77769  
 Sequencing vector: M13; L08821  
 Chemistry: Dye-terminator Big Dye; 96% of reads

Assembly program: Phrap; version 0.990329  
 Consensus quality: 247540 bases at least Q40  
 Consensus quality: 264763 bases at least Q30  
 Consensus quality: 273200 bases at least Q20  
 Estimated insert size: 327546; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-rip estimation  
 Quality coverage: 4.1x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
 \* NOTE: This sequence may represent more than one clone.  
 \* NOTE: This is a working draft sequence. It currently  
 \* consists of 28 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1  
 \* 65488: contig of 65488 bp in length  
 \* 65489: gap of unknown length  
 \* 65589: contig of 63497 bp in length  
 \* 129086: gap of unknown length  
 \* 129185: gap of unknown length  
 \* 129186: contig of 32432 bp in length  
 \* 161618: gap of unknown length  
 \* 161717: gap of unknown length  
 \* 171830: contig of 10113 bp in length  
 \* 171831: gap of unknown length  
 \* 171931: contig of 3165 bp in length  
 \* 175095: gap of unknown length  
 \* 175096: gap of unknown length  
 \* 175196: contig of 4010 bp in length  
 \* 179205: contig of 4010 bp in length  
 \* 179305: gap of unknown length  
 \* 179306: contig of 4983 bp in length  
 \* 184288: gap of unknown length  
 \* 184389: contig of 3895 bp in length  
 \* 188283: gap of unknown length  
 \* 188383: gap of unknown length  
 \* 188384: gap of unknown length  
 \* 192040: contig of 3657 bp in length  
 \* 192041: gap of unknown length  
 \* 192141: contig of 3001 bp in length  
 \* 195141: gap of unknown length  
 \* 195142: gap of unknown length  
 \* 195241: contig of 3114 bp in length  
 \* 198355: gap of unknown length  
 \* 198456: gap of unknown length  
 \* 198457: gap of unknown length  
 \* 201722: contig of 3267 bp in length  
 \* 201823: gap of unknown length  
 \* 205001: contig of 3179 bp in length  
 \* 205102: gap of unknown length  
 \* 205101: gap of unknown length  
 \* 207218: contig of 2017 bp in length  
 \* 207219: gap of unknown length  
 \* 209588: contig of 2640 bp in length  
 \* 209589: gap of unknown length  
 \* 209595: gap of unknown length  
 \* 212786: contig of 2828 bp in length  
 \* 212787: gap of unknown length  
 \* 212886: gap of unknown length  
 \* 215444: contig of 2558 bp in length  
 \* 215445: gap of unknown length  
 \* 215446: gap of unknown length  
 \* 215545: contig of 2035 bp in length  
 \* 217579: gap of unknown length  
 \* 217580: gap of unknown length  
 \* 220431: contig of 2752 bp in length  
 \* 220432: gap of unknown length  
 \* 220532: gap of unknown length  
 \* 222696: contig of 2165 bp in length  
 \* 222697: gap of unknown length  
 \* 222797: gap of unknown length  
 \* 225582: contig of 2786 bp in length  
 \* 225583: gap of unknown length  
 \* 227743: contig of 2061 bp in length  
 \* 227744: gap of unknown length  
 \* 227843: gap of unknown length  
 \* 230046: contig of 2203 bp in length  
 \* 230047: gap of unknown length  
 \* 230147: gap of unknown length  
 \* 232313: contig of 2167 bp in length  
 \* 232314: gap of unknown length  
 \* 232414: gap of unknown length  
 \* 234578: contig of 2165 bp in length  
 \* 234579: gap of unknown length  
 \* 234678: gap of unknown length  
 \* 237392: contig of 2614 bp in length  
 \* 237393: gap of unknown length  
 \* 237398: contig of 2596 bp in length  
 \* 239988: gap of unknown length  
 \* 240088: gap of unknown length

```

FEATURES      * 240089 242513: contig of 2425 bp in length.
SOURCE        1. 242513
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="Xp"
               /clone="RP11-386120"
BASE COUNT    71180 a 49220 c 48511 g 70846 t 2756 others
ORIGIN

Query Match      4.9%: Score 81.2; DB 2; Length 242513;
Best Local Similarity 49.0%: Pred. No. 0.015;
Matches 220; Conservative 0; Mismatches 228; Indels 1; Gaps 1;

QY 50 attgtgtcatttaagatttaagagatgcacaaatttgaagataattagatct 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 128435 ATATATATATAATTTNANNAGATTNTATATANAATAATTAATAATNTATTA 128494

QY 110 tttagtaaaagatttcttcttggaagcaatgaggggttcttcttggttttataag 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 128495 TTTTATATATATAATTAATAATTAATAATTTTAAATTTTATNTATTAATTAAT 128554

QY 170 tgaatcaagatttatacttcttctttaaagaatgagatttatacttcaag 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 128555 TATATTAATTAATTAATTAATTTTATATTTTAAATTTTAAATTTTAAATTTTAAAN 128614

QY 230 agaatttcatg-attctttaaataatgagatataagaatcttgcgttaaggtca 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 128615 AATATTAATAAATTTTAAATTAATAATTAATTAATTTTAAATTTTAAATTAATTTT 128674

QY 289 attcttctgctatgctttaaataaacaatgagcttacttctctgtaagataaaaa 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 128675 ATATATATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 128734

QY 349 agattatattcttcttggaaggtgatctttaaagaagcaatgataaattt 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 128735 ATNAATTTTATTTTAAATTTTAAATTAATTAATTAATTAATTTTATTTTATTTT 128794

QY 409 actaataaataaataaataaagcaaaatttataaagattatataaaggattt 468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 128795 AATTAATAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 128854

QY 469 ttacatgaaaggaagcctaaagattt 497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 128855 TTATTAATTAATTAATTAATTAATTAATTAATTT 128883

RESULT 13
CONSOLRGO 175229 bp DNA HTG 29-JUN-2001
LOCUS Homo sapiens chromosome 14 clone R-828K24, *** SEQUENCING IN
DEFINITION PROGRESS ***, 2 ordered pieces.
ACCESSION AL159141
VERSION AL159141.4 GI:14588844
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
1 (bases 1 to 175229)
Helliq.R., Petit.J.L., Vico.V., Dasilva.C., Robert.C., Mincker.P.,
Brottier.P., Catolico.L., Barbe.V., Pelletier.E., Artigaudave.F.,
Ley,M., Eckenberg,R., Bruls,T., deBeraudins,V., Cruaud,C.,
Gyapay,G., Saurin,W. and Weissensbach,J.
Sequencing of the human chromosome 14
Unpublished
2 (bases 1 to 175229)
Genoscope.
Direct Submission
Submitted (29-JUN-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

```

```

COMMENT
On Jul 3, 2001 this sequence version replaced gt:10944754.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: Sequef@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc. . . even if efforts are made to eliminate these
contaminating sequences.
Upstream BAC (overlapping the T7 end) : R-618B21
Downstream BAC (overlapping the SP6 end) : R-807G16 (AC-AL136332)
----- Summary Statistics -----
Assembly program: Phrap: version 2.0
Quality coverage: 6.28x in 920 bases; sum-of-contigs
-----
Contigs composition :
71802 bp contig from 1 to 71802
103327 bp contig from 71903 to 175229
-----
Overall quality chart :
Range : bases
0 : 363
1 - 9 : 12
10 - 19 : 99
20 - 29 : 216
30 - 39 : 714
40 - 49 : 2946
50 - 59 : 6725
60 - 69 : 13088
70 - 79 : 33326
80 - 89 : 64652
90 - 99 : 53078
-----
Percentage of bases with a quality value >= 40 : 99 %.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 71802: contig of 71802 bp in length
* 71803 71902: gap of 100 bp
* 71903 175229: contig of 103327 bp in length.
Location/Qualifiers
1. 175229
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone_11b="R-828K24"
/clone_11b="RP11-11"
18301..18385
/notes="matching EMBL:M79260
Rdbb:RH28481
Rdbb:RH28423
dbSTS:STS9346
Identified using the e-PCR software (G. Schuler)"
110596..110728
/notes="matching EMBL:M52863
Rdbb:RH45925
dbSTS:STS38993
Identified using the e-PCR software (G. Schuler)"
BASE COUNT 44138 a 41136 c 42653 g 47199 t 103 others
ORIGIN

```













GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2002, 23:30:59 ; Search time 311.54 Seconds  
(without alignments)  
40.132 Million cell updates/sec

Title: US-09-004-395-2

Perfect score: 1719  
Sequence: 1 MKRKAKILFLLSTVLPAQ.....FQNFVERIESKPESSPKN 341

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1701	99.0	341	1	FLAA_BORBU
2	403	23.4	337	1	FLAA_BORBU
3	361.5	21.0	350	1	FLAA_TREPA
4	331.5	19.3	320	1	FLAA_TREPA
5	109	8.3	444	1	YCD_BACSU
6	108	6.3	580	1	Y686_METUA
7	106.5	6.2	695	1	VAT1_METUA
8	105.5	6.1	705	1	CMBA_BACSU
9	105	6.1	902	1	ITR1_PIG
10	104	6.1	499	1	XIIB_BACSU
11	103	6.0	650	1	ITR1_LYCES
12	102.5	6.0	914	1	ITR1_MESAU
13	101.5	5.9	613	1	CGAA_CLOBI
14	101.5	5.9	622	1	YF54_METUA
15	101.5	5.9	676	1	HS7C_TRYRB
16	101	5.9	1682	1	MSR1_PLAF3
17	100	5.8	701	1	SYGB_HELPJ
18	99.5	5.8	907	1	ITR1_MOUSE
19	98.5	5.7	416	1	P46_MYCH
20	98.5	5.7	644	1	HS72_LYCES
21	98	5.7	707	1	YJ9C_YEAST
22	97.5	5.7	721	1	ENP1_TORCA
23	96	5.6	1875	1	MLP1_YEAST
24	95.5	5.6	434	1	ENO1_LYCE
25	95.5	5.6	931	1	HS71_ARATH
26	95.5	5.6	931	1	DAP1_YEAST
27	95.5	5.6	1104	1	BUD2_YEAST
28	95	5.5	641	1	HS74_YEAST
29	95	5.5	1024	1	SYIP_STRAU
30	95	5.5	1125	1	MPD_BORBU
31	95	5.5	1701	1	MSR1_PLAF3
32	95	5.5	1701	1	MSR1_PLAF3
33	94.5	5.5	680	1	OPDA_ECOLI

34	94	5.5	655	1	HS70_DAUCA	P26791	daucus caro
35	93.5	5.4	406	1	RENT_HUMAN	P00797	homo sapien
36	93.5	5.4	494	1	AMY1_SACFI	P21567	saccharomyc
37	93.5	5.4	651	1	HS7C_PETHY	P09189	petunia hyb
38	93.5	5.4	1672	1	PMR1_CHUMU	O9P132	chlamydia m
39	93	5.4	282	1	RUIA_HUMAN	P09012	homo sapien
40	93	5.4	282	1	RUIA_XENLA	P45429	xenopus lae
41	93	5.4	341	1	MDHM_BRANA	O43744	brassica na
42	93	5.4	454	1	MDHM_AQUAE	O67373	aquifex neo
43	93	5.4	608	1	KU70_HUMAN	P12956	homo sapien
44	93	5.4	701	1	SYGB_HELPJ	P56454	helicobacte
45	93	5.4	1296	1	ASAI_ENTFA	P17953	enterococcu

## ALIGNMENTS

RESULT 1  
ID FLAA\_BORBU STANDARD; PRT; 341 AA.  
AC P70856; O51612; Q44876;  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE FLAGELLAR FILAMENT OUTER LAYER PROTEIN PRECURSOR (SHEATH PROTEIN).  
GN FLAA OR BB0668.  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_Taxid=139;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-212;  
RX MEDLINE-97144545; PubMed-8990312;  
RA Ge Y., Charon N.;  
RT "An unexpected flaa homolog is present and expressed in Borrelia burgdorferi."  
RT J. Bacteriol. 179:552-556(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 35210 / B31;  
RX MEDLINE-98065943; PubMed-9403685;  
RA Fraser C.M., Casjens S., Huang W.M., Hickey E.K., Gwinn M., Lathigra R., White O., Fleischmann R.D., Richardson D., Dougherty B., Tomb J.F., Fleischmann R.D., Salzberg S., Hanson M., van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J., Utecht T., White L., McDonald L., Arlath P., Bowman C., Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B., Smith H.O., Venter J.C.;  
RT "Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi."  
RT Nature 390:580-586(1997).  
RN [3]  
RP SEQUENCE OF 276-341 FROM N.A.  
RC STRAIN-212;  
RX MEDLINE-9848936; PubMed-9765799;  
RA Old I.G., Trueba G.A., Saint-Girons I., Johnson R.C.;  
RT "A chea chew operon in Borrelia burgdorferi, the agent of Lyme disease."  
RL Res. Microbiol. 148:191-200(1997).  
CC -!- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.  
CC -!- SUBUNIT: THE FLAGELLUM CONSISTS OF AN OUTER LAYER COMPOSED OF REPEATING UNITS OF FLAA AROUND A CORE THAT CONTAINS SEVERAL ANTIGENICALLY RELATED POLYPEPTIDES.  
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC FLAGELLUM.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL: U62900; AAC44770.1;
DR EMBL: AE001168; AAC67025.1; ALT_INIT.
DR EMBL: X91907; CAA63001.1;
DR TIGR: B06668;
KW flagella; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 341
FT CONFLICT 254 255 EQ -> VK (IN REF. 1).
FT CONFLICT 317 318 MP -> IA (IN REF. 1 AND 3).
SQ SEQUENCE 341 AA: 38441 MW: 60814758BADF3451 CRC64.

Query Match 99.0%; Score 1701; DB 1; Length 341;
Best Local Similarity 98.8%; Pred. No. 3.5e-117;
Matches 337; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRKAKSLIFLLSTVLVAQETDGLAEGSKRAPEGLVDFEALRDPSTRLDTNVD 60
DB 1 MKRKAKSLIFLLSTVLVAQETDGLAEGSKRAPEGLVDFEALRDPSTRLDTNVD 60
QY 61 YVSGASGIVPEDEWVDLGINNSVLLTPSARLQAYVKNVAVPAVYKSESKRYAGDTI 120
DB 61 YVSGASGIVPEDEWVDLGINNSVLLTPSARLQAYVKNVAVPAVYKSESKRYAGDTI 120
QY 121 LGVRLFPSSYSSSAMIMPPFKIPYSGESGNOFLGKLIDIKMKELIKSVYSLGVEI 180
DB 121 LGVRLFPSSYSSSAMIMPPFKIPYSGESGNOFLGKLIDIKMKELIKSVYSLGVEI 180
QY 181 DLEVLFEDMGMEYAVSMGTLKFKGMADLINSNPYIPNISRIIKDVPNPPLASSKMR 240
DB 181 DLEVLFEDMGMEYAVSMGTLKFKGMADLINSNPYIPNISRIIKDVPNPPLASSKMR 240
QY 241 FKAFVSKSHSKVKNFFYYKDLVLDKLSVSDSDSESVKVEYETSGTESLRLK 300
DB 241 FKAFVSKSHSKVKNFFYYKDLVLDKLSVSDSDSESVKVEYETSGTESLRLK 300
QY 301 AHETFRVLKLRKISIAEGSFONFEKIESEKPESSPKN 341
DB 301 AHETFRVLKLRKISIAEGSFONFEKIESEKPESSPKN 341

RESULT 2
FLAA_SPIAU STANDARD; PRT; 337 AA.
AC P21982;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FLAGELLAR FILAMENT OUTER LAYER PROTEIN PRECURSOR (SHEATH PROTEIN).
GN FLAA.
OS Spirochaeta aurantia.
OC Bacteria; Spirochaetales; Spirochaetaceae; Spirochaeta.
OX NCBI_TaxID=147;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89155480; PubMed=2921247;
RA Brahamsa B., Greenberg E.P.;
RT "Cloning and sequence analysis of flaa, a gene encoding a Spirochaeta
  aurantia flagellar filament surface antigen.";
RL J. Bacteriol. 171:1692-1697(1989).
RN [2]
RP REVISIONS.
RA Greenberg E.P.;
RL Submitted (JAN-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 22-49.
RC STRAIN=M1;
RX MEDLINE=91123217; PubMed=1991729;
RA Parales J. Jr., Greenberg E.P.;
RT "N-terminal amino acid sequences and amino acid compositions of the
  Spirochaeta aurantia flagellar filament polypeptides.";
RL J. Bacteriol. 173:1357-1359(1991).

```

```

CC -----
CC -1- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.
CC -1- SUBUNIT: THE FLAGELLUM CONSISTS OF AN OUTER LAYER COMPOSED OF
CC REPEATING UNITS OF FLAA AROUND A CORE THAT CONTAINS ONE OR ALL OF
CC FIVE ANTIGENICALLY RELATED POLYPEPTIDES.
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC FLAGELLUM.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M24459; AAA9273.1;
DR PIR: A32814; A32814.
KW flagella; Periplasmic; Signal.
FT SIGNAL 1 21
FT CHAIN 22 337
SQ SEQUENCE 337 AA: 36857 MW: 332648096785635 CRC64;

Query Match 23.4%; Score 403; DB 1; Length 337;
Best Local Similarity 30.4%; Pred. NO. 2e-22;
Matches 100; Conservative 67; Mismatches 14; Indels 38; Gaps 9;

QY 1 MKRKAKSLIFLLSTVLVAQETDGLAEGSKRAPEGLVDFEALRDPSTRLDTNVD 60
DB 1 MKRKAKSLIFLLSTVLVAQETDGLAEGSKRAPEGLVDFEALRDPSTRLDTNVD 60
QY 61 YVSGASGIVPEDEWVDLGINNSVLLTPSARLQAYVKNVAVPAVYKSESKRYAG 117
DB 61 YVSGASGIVPEDEWVDLGINNSVLLTPSARLQAYVKNVAVPAVYKSESKRYAG 117
QY 118 DTLGVRVLFPSYS-QSSAMIMPPFKIPY-----SGESGNOFLGKLIDIKMK 167
DB 118 DTLGVRVLFPSYS-QSSAMIMPPFKIPY-----SGESGNOFLGKLIDIKMK 167
QY 167 STQINILGRNYINRLSLLEQNGDERELVQYLFDMKSLQWNNPNYQYEVRRRDQ- 225
DB 167 STQINILGRNYINRLSLLEQNGDERELVQYLFDMKSLQWNNPNYQYEVRRRDQ- 225
QY 228 DVPNPPLASSKMRFAFRVSKSHSKVKNFFYYKDLVLDKLSVSDSDSESV- 284
DB 228 DVPNPPLASSKMRFAFRVSKSHSKVKNFFYYKDLVLDKLSVSDSDSESV- 284
QY 285 -----FKVYETSGTESLRLKKAHETFK 306
DB 285 -----FKVYETSGTESLRLKKAHETFK 306
QY 306 LKREDEYRNFELAKGLNQLVRSLEKRR 314
DB 306 LKREDEYRNFELAKGLNQLVRSLEKRR 314

RESULT 3
FLAA_TREPA STANDARD; PRT; 350 AA.
AC P18193;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FLAGELLAR FILAMENT OUTER LAYER PROTEIN PRECURSOR (SHEATH PROTEIN).
GN FLAA OR TP0249.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90307197; PubMed=2194955;
RA Isaacs R.D., Radolf J.D.;
RT "Expression in Escherichia coli of the 37-kilodalton endoflagellar
  sheath protein of Treponema pallidum by use of the polymerase chain
  reaction and a T7 expression system.";
RL Infect. Immun. 58:2025-2034(1990).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RA STRAIN-NICHOLS:
RX MEDLINE=98332770; PubMed=9658976;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Kethum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Kholak H., Richardson D., Howell J.R., Chidambaram M., Utterback T.,
RA McDonald L., Artchak P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388(1998).
RN 131
RP SEQUENCE OF 30-350 FROM N.A.
RC STRAIN-NICHOLS,
RX MEDLINE=90035409; PubMed=2680972;
RA Isacs R.D., Hanke J.H., Guzman-Verduzco L.-M., Newport G.,
RA Agabian N., Norgard M.V., Lukehart S.A., Redolf J.D.;
RT "Molecular cloning and DNA sequence analysis of the 37-kilodalton
RT endoflagellar sheath protein gene of Treponema pallidum.";
RL Infect. Immun. 57:3403-3411(1989).
CC -1- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.
CC -1- SUBUNIT: THE FLAGELLUM CONSISTS OF AN OUTER LAYER COMPOSED OF
CC REPEATING UNITS OF FLAA AROUND A CORE THAT CONTAINS SEVERAL
CC ANTIGENICALLY RELATED POLYPEPTIDES.
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC FLAGELLUM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63142; AAA27477.1; .
DR EMBL; AE001206; AAC65235.1; .
DR EMBL; M26525; AAA27476.1; .
DR TIGR; TP0249; .
KW Flagella; Periplasmic; Signal; Complete proteome.
FT SIGNAL
FT CHAIN 21 350 POTENTIAL FLAELLAR FILAMENT OUTER LAYER PROTEIN.
SQ SEQUENCE 350 AA; 38857 MW; 2BC7CFEFA0CDBA67 CRC64;
-----
QY Query Match 21.08; Score 361.5; DB 1; Length 350;
QY Best Local Similarity 28.58; Pred. No. 2.2e-19;
QY Matches 100; Conservative 66; Mismatches 120; Indels 65; Gaps 11;
-----
QY 6 KSIL--FLLSTVLPAGCTDGLAEGRKAEPGELVDFEAELARDPSSRLDITNVDDVY 63
QY :|:| | : : : :|:|:| | | | | | | | | | | | | | | | | |
QY 22 ESVLIDFPLKLNADIDAKSGGTHRR-----TYLDAISLND-----TSTTD-- 63
QY :|:| | : : : :|:|:| | | | | | | | | | | | | | | | | |
QY 64 SGASGIYKPEDMYVDLGINNMSVLLTPSARLQAAVYKNSVPAVYKSSSKRYAGDTILGV 123
QY :|:| | : : : :|:|:| | | | | | | | | | | | | | | | | |
QY 64 -----EQKALMRSSSLAAVQMEVYVNSARNPVAAHARSVTEAPVSEGRKSPAGERVLGV 117
QY :|:| | : : : :|:|:| | | | | | | | | | | | | | | | | |
QY 124 RVLFPSTY--SQSSAMIMPFKTIPTFS-----GEGSN-----QL--GKGLIDN 162
QY :|:| | : : : :|:|:| | | | | | | | | | | | | | | | | |
QY 118 RLLPPTWDSNANMKIPAFVIPAPEVMAQVDDGVQVQAPTEERKSGKGRFEDGQVYKN 177
QY :|:| | : : : :|:|:| | | | | | | | | | | | | | | | | |
QY 163 IKTKMEIKVSYSLGLEYLDFEDNMGMEAYSMGLTKFKFGMADLMSNNTYIPNTSS 222
QY :|:| | : : : :|:|:| | | | | | | | | | | | | | | | | |
QY 178 VGVLSKINVTNYGNNYPIGLLYVMARDQGEVHRYRFGILLDSKMKELVWNNPSTISDVS 237
QY :|:| | : : : :|:|:| | | | | | | | | | | | | | | | | |
QY 223 RIKDDVNPYPLASSKMKFAFRVSKSHSSKAVKNFIYVKDLRVLYDKLSVSDIDSE 282
QY :|:| | : : : :|:|:| | | | | | | | | | | | | | | | | |
QY 238 REVR-LYEVYPASTRPHVYVEGFMVTRDAAHAGGDVYGFKDKVITIDAVLSTVDFADE 296
QY :|:| | : : : :|:|:| | | | | | | | | | | | | | | | | |
QY 283 SVPKYEDISGSESRKLKAHETFKRVLRLREKISIAESDFQNFVYKISEK 333
QY :|:| | : : : :|:|:| | | | | | | | | | | | | | | | | |
QY 297 DLW-----GIQARREAE-----RRVEYARFGQOOVLVYIQEK 330
QY :|:| | : : : :|:|:| | | | | | | | | | | | | | | | | |

```

```

RESULT 4
FLAI_TREH4
ID FLAI_TREH STANDARD: PRT: 320 AA.
AC P32520; P80157;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FLAGELLAR FILAMENT OUTER LAYER PROTEIN FLAI1 PRECURSOR (44 KDA SHEATH
DE PROTEIN) FLAI.
GN FLAI1 OR FLAI.
OS Treponema hyodysenteriae (Serpulina hyodysenteriae).
OC Bacteria; Spirochaetales; Brachyspiraceae; Brachyspira.
OX NCBI_TaxId:159;
[1]
RN RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-38.
RC STRAIN-C5;
RX MEDLINE=92307926; PubMed=1612759;
RA Koopman M.B.H., de Leeuw O.S., van der Zeijst B.A.M., Kusters J.G.;
RA van der Zeijst B.A.M., Kusters J.G.;
RA "The periplasmic flagella of Serpulina (Treponema) hyodysenteriae are
RA composed of two sheath proteins and three core proteins."
RA J. Gen. Microbiol. 138:2697-2706(1992).
CC -1- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.
CC -1- SUBUNIT: THE FLAGELLUM CONSISTS OF AN OUTER LAYER COMPOSED OF TWO
CC SHEATH PROTEINS, FLAI1 (44 KDA) AND FLAI2 (35 KDA) AROUND A CORE
CC THAT CONTAINS THREE PROTEINS FLAI3 (37 KDA), FLAI2 (34 KDA) AND
CC FLAI3 (32 KDA).
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC FLAGELLUM.
CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC CC DR EMBL; X63006; CAA44735.1; -.
CC DR EMBL; A24079; CAA01715.1; -.
CC DR PIR; A43824; A43824.
CC KW Flagella; Signal; Periplasmic.
CC FT SIGNAL 1 19
CC FT CHAIN 20 320
CC FT FLAI1.
CC FT FLAGELLAR FILAMENT OUTER LAYER PROTEIN
CC FT
CC FT CONFLICT 21 21 T -> S (IN REF. 2).
CC FT SEQUENCE 320 AA; 35986 MW; 3CF8C27B07F8646B4 CRC64;
SQ
Query Match 19.3%, Score 331.5; DB 1; Length 320;
Best Local Similarity 28.8%; Pred. No. 3e-17;
Matches 90; Conservative 63; Mismatches 129; Indels 31; Gaps 8;
OY 9 LEFLLSTVLFQGEIDGLAGSKRAPEGLYDFEALARDPSSRLDITNRYDVYVSGASG 68
DB 4 LEVVLSTLIFIAASAYGLTNSI-----LDFA-----LTGMADNLDAGEED 43
OY 69 IVKPEDMVVDGLINNMVSULLTPSARLQAYVKNVSAVPAVKSSEKRYAGDTLLGVRLFP 128
DB 44 TNEVPAVENLNDMMVYWLNESARLTFEBRRNSYTVTNDKSCNGAMEAGKVLGVRIHP 103
OY 129 SYSQSS-AMINPEPFIPIYSGESGQFL-GKGLDINIKTKIKSVYSLGEIDLEVLF 186
DB 104 LAANMSVALVRYEELWEGADGTRKYTEGKQVHNHVEIKISISSMYGRANLLISFYNL 163

```

```

OY 187 EGMNMEVYVSGMGLTKFKGMADLISNPNYINISRIITKDDVNPVPLASSCMRKAAPV 246
DB 164 QNEFELKSPYMGVYFYFNKMRQVRENREKRYLPLVNDVRLVIRE-PLYPRAIPSPKIDSLGF 222
OY 247 SKSHSSKRVKNLFYVKDLRVLYDKLSVSDSIDSESVFKYVETSGTESLRKLKAHEYFK 306
DB 223 YRTKPKTKDEFTTYVKKVDYLVLEVVVVVDFEEDIDDEATWQLTK--TENDRK-QAIES- 276
OY 307 RVLRLEKRSIAIE 319
DB 277 --ARIREQAELRD 287

RESULT 5
YCXD_YACSU STANDARD: PRT: 444 AA.
AC 008792:
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 50.8 KDA PROTEIN IN SRFA4-SFP INTERGENIC REGION (ORF8).
CN YCXD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1423;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=93360813; PubMed=8355609;
RA Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
RA Venema G., van Sinderen D.,
RT "Sequence and analysis of the genetic locus responsible for surfactin
RT synthesis in Bacillus subtilis."
RL Mol. Microbiol. 8:821-831(1993).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN=168.
RC MEDLINE=97124189; PubMed=89659502;
RX Yamane K., Kumano M., Kurita K.;
RA "The 23 degrees-36 degrees region of the Bacillus subtilis chromosome:
RT determination of the sequence of a 146 kb segment and identification
RT of 113 genes."
RL Microbiology 142:3047-3056(1996).
CC -I- COFACTOR: PYRIDOXAL PHOSPHATE (POTENTIAL).
CC -I- SIMILARITY: IN THE C-TERMINAL, TO AMINOTRANSFERASIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X70356; CAA49821.1; -
DR EMBL, D50453; BAA08990.1; -
DR EMBL, Z99105; CAB12150.1; -
DR EMBL, Z99106; CAB12164.1; -
DR Subtilist; BG10175; YCXD.
DR InterPro; IPR000524; HTH_Gntr.
DR Pfam; PF00392; gntr; 1.
DR PRINTS; PR000035; HTHGNTR.
DR SMART; SM00345; HTH_GNTR; 1.
KM Hypothetical protein: Transferase; Amino transferase;
KW Pyridoxal phosphate: Complete proteome.
FT BINDING 286 286 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 444 AA: 50755 MW; 1C392C3D386696E7 CRC64;

```

Query Match	6.38;	Score 109;	DB 1;	Length 444;
Best Local Similarity	21.68;	Pred. No. 0.83;		

Matches 75; Conservative 42; Mismatches 126; Indels 104; Gaps 15;

```

QY      29 SKRPEGLVYDFEFLKRDSESTLDTLNVYD-----VYSGASGt---69
Db      73 SKSGQPGP--IDFATSAEDP-----DVFPYLDFOHCINKAIDTYKNDLFTYGPBKGLPSL 125
QY      70 -----VKEDMWVDYDGINNWSVL-----TPSARLQAVKN 100
Db      126 IRLVRLKLLATQVADERHNIPTTSGVOALSLLCAMPFPGKKEKIALIOPGYHLMWOLE 185
QY      101 SVVAPAVVSKSKRYACDITLGVRLVLPFSPYSQSSAMITPPKIPFYSGESGNOLFCKGLI 160
Db      186 TLGLPAILGVKRTTE--GIDIAKIVRLTQTESIKFEPYMPRFHNp-----LGSLS 233
QY      161 DNITKMKIKSVYSLGTEIDLEVLFDPMGMMEVYASMGTLKFKGADLIMSNPTIPNI 220
Db      234 ERDKO-----ELVRLAEDVYILVEDDY-----LGLDEENKKAD--PLAYDL 274
QY      221 SSRITKDDVPNYPLASSKMRKAEVRSKSHSKYKNFIYVKDLRVLYDKLVSIDSID 280
Db      275 SSHYI-----YLSKFSKMMPPGLRYCAAAVLPALRLDTFYA-----YKXLN-DIDCSMI 321
QY      281 SESVFKVYETSGTESLRKLKNAHEFRK-RVLYLREKI-----SIAGSF 332
Db      332 SQAMLEITKSGMYGRKREKIRDSYKERSLRHAIITRHQJGSGRF 368

```

RESULT	6
ID	Y686_METUA
AC	Q58099;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	HYPOTHETICAL PROTEIN MJ0686.
GN	MJ0686.
OS	Methanococcus jannaschii.
OC	Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC	Methanococcus.
OX	NCBI_TaxID=2190;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX	MEDLINE=96337999; PubMed=8688087;
RA	Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA	Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA	Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA	Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA	Scott J.L., Geoghegan N.S.M., Meldrum J.F., Fuhrmann J.L., Nguyen D.,
RA	Uutterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA	Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA	Klank H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT	"Complete genome sequence of the methanogenic archaeon, Methanococcus
RL	jannaschii".
CC	Science 273:1058-1073(1996).
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on ways
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/1sb.ch">http://www.isb-sib.ch/announce/1sb.ch</a> ).
CC	or send an email to <a href="mailto:license@1sb.ch">license@1sb.ch</a> ).
DR	EMBL: U67515; AAB98681.1; -.
DR	TIGR: MJ0686; -.
DR	Hypothetical protein: Complete proteome.
QW	SEQUENCE 580 AA; 68392 MW; C33CAL46575B2506 CRC64.

Query Match	6.38;	Score 108;	DB 1;	Length 580;
Best Local Similarity	21.78;	Pred. No. 1.4;		
Matches	56;	Conservative	38;	Mismatches 66;
				Indels 98;
				Gaps 11.

```

QY 112 SKRAGDTILGVRLFPSSQSSAMIPPKIPYSGSGNOFLGKLDINIKTKREIKV 171
DB 376 SERISGST-----ESKMKDKREVSQWIENNKKI--PEVEVLKI--EIQY 417
QY 172 SVSLGVEIDLEVLFEEDMNGMEYAYSMGLKFGKMDLIWSNPV---YINISSRIKID 228
DB 418 GAVSL-----KKVSSGSSKRYKGVNLIKRNKPMIDYKP-----DN 453
QY 229 VPNYPLASSMKRFKAP-----RVSKSHSSKRYKNFIYVK 262
DB 454 IAYKLEDDHHIFPKGFLRNKGISNEYIDSVLNKTPILDETNNKISKSPSK-----YVK 507
QY 263 DLAVLDKLSVSDSD-----SESVFKVYETSGTESLRKLKAHEFRKRLVL 311
DB 508 EMEIKRNKGLSDEAVNKKVEILKGHINDEMEFILLRNT--DLSLKDEIEENFRIEL 566
QY 312 REKISIAEGSFQNFVEKI 329
DB 567 REKL-----ILEKI 575

RESULT 7
VATILMETUA STANDARD: PRT: 695 AA.
AC 057675;
DB 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
V-TYPE ATP SYNTHASE SUBUNIT I (EC 3.6.1.34) (V-TYPE ATPASE SUBUNIT I).
GN ATP1 OR MJ0222.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1/DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sulton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurt M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii."
RL Science 273:1058-1073(1996).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67478; AAB98208.1; -.
CC DR TIGR: M02222; -.
CC DR InterPro: IPR002490; V_ATPase_sub_a.
CC DR Pfam: PF01496; V_ATPase_sub.a; 1.
CC KM Hydrolase: Hydrogen ion transport; Transmembrane; Complete proteome.
CC FT TRANSMEM 392 412 POTENTIAL.
CC FT TRANSMEM 425 445 POTENTIAL.
CC FT TRANSMEM 497 517 POTENTIAL.
CC FT TRANSMEM 534 554 POTENTIAL.
CC FT TRANSMEM 556 576 POTENTIAL.
CC FT TRANSMEM 605 625 POTENTIAL.

```

```

FT TRANSMEM 627 647 POTENTIAL.
SQ SEQUENCE 695 AA; 76953 MW; F03E3CEBE29D53D CRC64;

Query Match
Best Local Similarity 23.5%; Score 106.5; DB 1; Length 695;
Matches 84; Conservative 41; Mismatches 108; Indels 125; Gaps 19;

QY 68 GIVPEDMVVDLGINNSVLLTPSARLQAYKNSVAPAVKSSKRYAGDTILGVRLV 126
DB 32 GIVELCDLSEKLEDELEKTLSPSSAD-YVRN--VYSLMK-----AG-----RLD 76
QY 127 -FSPYSQSSAMI-----MPPKIPYSGSGNOFLGKLDINIKTKREIKV 161
DB 77 MESSVSQKETSIDKILNPKPKKVSFNSYQEVYDAEKVLNISKVEVDGPAERLSELD 136
QY 162 NIKT-MKEIKVSV-YSLGVEIDLEVLFEEDMNGMEYAY-----SMGLTKFGKMDLI 210
DB 137 NKKSKLQKQKQISYLGKLEPDLKYL-----GSGEVFVIGASVPKREKGLK-----AELD 188
QY 211 WSNPNYIPNIS-SRIKDVNPVPLASSMKRFKAPRVSKSHSSKRYKNFIYVKDLRVLYD 269
DB 189 KVADGYIGIFSGSEFEKDKIRVPIVPTLKEKLENV-----LSEIRKFEFERYDI----- 239
QY 270 KLSVSDSDSD-----ESVFKVYETSGTESLRKL--AHETFRVLKREKISI- 317
DB 240 -----SDVEGTPSEALSKIESLKAIESRNSLIEKRLAKLAKWEKELLAVELLSTE 292
QY 318 -----AEGSFQNFVEKI-----EVEKIESKPESSP 339
DB 293 KARGDAVSQFGKTRTYIEMVAPARDAEKAKSLIENSADGFARVEITPEPEPEKIP 350

RESULT 8
CMBB_BACSU STANDARD: PRT: 705 AA.
AC 002113;
DB 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
V-TYPE ENHANCER PRECURSOR (MODIFIER PROTEIN OF MAJOR AUTOLYSIN).
GN LYT8 OR CMBB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93018996; PubMed=1357079;
RA Lazarevic V., Margot P., Soldo B., Karamata D.;
RT "Sequencing and analysis of the Bacillus subtilis lyt8BAC divergon: a
RT regulatory unit encompassing the structural genes of the N-
RT acetylmutamoyl-L-alanine amidase and its modifier."
RL J. Gen. Microbiol. 138:1949-1961(1992).
CC [2]
CC SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
CC RP MEDLINE=92407479; PubMed=1356138;
CC RX Kuroda A., Rashid H.M., Sekiguchi J.;
CC "Molecular cloning and sequencing of the upstream region of the major
CC Bacillus subtilis autolysin gene: a modifier protein exhibiting
CC sequence homology to the major autolysin and the spoliid product."
CC J. Gen. Microbiol. 138:1067-1076(1992).
CC -1- FUNCTION: POSSIBLY INVOLVED IN CELL WALL METABOLISM DURING SPORE
CC FORMATION. ENHANCES THE AMIDASE ACTIVITY APPROXIMATELY THREEFOLD.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -1- SIMILARITY: WITH THE N-TERMINAL OF LYT8 AT ITS N-TERMINAL AND THE
CC B. SUBTILIS SPOIID GENE AT ITS C-TERMINAL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```





Oy 175 SLGSEIDLEVLFEEDNOME-----YASMGILKRRKGMAADLIWSNPNTIPIISSILIK 226  
 Db 424 NLGCGHDVENNFLEVALLENGRAQRIEEDHDSAOQLQGYYDOV-ANP-----LTK 473  
 Oy 227 DDVENYR-----LASSKARFKAFR-----YKSHSSKSYKNFIFYVADLVLDKLSYSI 275  
 Db 474 DVEIQYPADAVLALTOHHNQYEGSEITVAAGRIADKLISSF-----KADYQA 521  
 Oy 276 DSD-----IDSESVFKVYETSG-----TESLRKLKAHEFKRYLKLREKISIAE 319  
 Db 522 SGDDGFTTCLVDEDEEKKKLLOEGRGHMLENVERLVAAYLTIOELAKRMKLEWAE 576

RESULT	10				
XYLB_BACSU	STANDARD:	PRT:	499 AA.		
ID	XYLB_BACSU				
AC	P39211;				
DT	01-FEB-1995 (Rel. 31, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	XYLUOSE KINASE (EC 2.7.1.17) (XYLUOKINASE).				
OS	XYLB.				
CS	Bacillus subtilis.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OC	Bacillus/Staphylococcus group; Bacillus.				
OX	NCBI_TaxID=1423;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Bocheret S., Klein C., Plkxa B., Hammelmann M., Ettlin K.D.;				
RT	"Sequencing of a 26 kb region of the Bacillus subtilis genome downstream of spoVJ.";				
RL	submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE OF 1-8 FROM N.A.				
RX	MEDLINE=65297769; PubMed=2994009;				
RA	Wilhelm M., Hollenberg C.P.;				
RT	"Nucleotide sequence of the Bacillus subtilis xylose isomerase gene: extensive homology between the Bacillus and Escherichia coli enzyme.";				
RL	Nucleic Acids Res. 13:5717-5722(1985).				
CC	-1- CATALYTIC ACTIVITY: ATP + D-XYLUOSE - ADP + D-XYLUOSE				
CC	5-PHOSPHATE.				
CC	-1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCOKINASE /				
CC	GLYCEROKINASE / XYLUOKINASE FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announcement/</a>				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; U66480; AAB41094.1; -				
DR	EMBL; X02795; CAA26563.1; -				
DR	EMBL; Z59113; CAB13645.1; -				
DR	Subtilist; BG10807; xyLB.				
DR	InterPro: IPR005577; FGGY_Kin.				
DR	Pfam: PF00370; FGGY_1.				
DR	PROSITE: PS00445; FGGY_KINASES_2; 1.				
DR	PROSITE: PS00933; FGGY_KINASES_1; 1.				
DR	Transferrase; Kinase; Xylose metabolism; Complete proteome.				
QO	SEQUENCE 499 AA; 55417 MW; 0772B7F67015B6E CRC64;				

Query Match	6.1%	Score 104	DB 1	Length 459
Best Local Similarity	20.9%	Pred. No. 2,2		
Matches	68	Conservative	57	Mismatches 133; Indels 66; Gaps 14.
QY	30	KRAEGELVDEAEALRDPSSRDLINVDYVYGASG----	IVKPE---	DMVVDLGI 81
DB	141	KEHEE-ELEKKTAVLRLKPDYVRFPMGTGLIHREYSDAAGTLLHTRKMSNDICNOI		199

```

Oy  82  NNM5VLLLP5ARLQAYVANSVVAAPAAVYSESEKRRVACGDTILGVRLLPFPYSOSSAMIPF 1411
Db  200  -----SADICPPLESMD-----CVGSLLRPHVAKTILLE--- 2230
Oy  142  KIPYSGESGNO-FLGKGLIDNIKTMYEKIVSVYSLGEYIDLE-----VLEEDMGME 193
Db  230  KTKVYAGADNACGAIGAGILSGSCITCSIGTSVILSEEEKRDEGKXHFPHGCKD 2893
Oy  194  YAYSMTGLKFKGMADLINSNPYIPNISRIKNDYPRVYPLASSKMRPKARVSEK--SHS 255
Db  230  SFYTMGVTLAAGYS-LDMFKRTFANSEEBOLLOGEAIPILGANGLLTPPLVGEKRPFA 3486
Oy  252  -SKVNFIFYVKD-----LRVLVDKLSVSDSIDD-----SESEVRYETSG--TES 295
Db  349  DSSINGSLIGMDGAHNKRKHPRLAIWEGITFSLHSEITELFREAGKSVHYVSVIGGAKNDT 408
Oy  296  LRKLKAHTEFKRVLLKREKISIAEGS 321
Db  409  WLQMODADLFENTRVILKLENEGPRAMGA 434

```

```

RESULT 11
H57L_LYCES
ID H57L_LYCES STANDARD: PRT: 650 AA.
AC P24629;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HEAT SHOCK COGNATE 70 KDA PROTEIN 1.
CN HSC-I.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
CX NCBI_Taxid=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV_VF36; TISSUE=Plst11;
RA MEDLINE=91370894; PubMed=1893113;
RT Lin T.Y., Duck N.B., Winter J., Folk W.R.;
RL "Sequences of two hac 70 cDNAs from Lycopersicon esculentum.";
Plant Mol. Biol. 16:475-478(1991).
CC -I- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X54029; CAA37970.1; -.
DR PIR: S14949; S14949.
DR HSSP: P19120; INGI.
DR InterPro: IPR001023; HSP70.
DR Pfam: PF00012; HSP70.1.
DR PRINTS: PR00301; HEATSHOCK70.
DR PROSITE: PS00297; HSP70.1; 1.
DR PROSITE: PS00329; HSP70.2; 1.
DR PROSITE: PS01036; HSP70_3; 1.
DR ATP-binding; Heat shock; Multigene family.
SQ SEQUENCE 650 AA: 71287 MW: 6C12ED08A7C8BD239 CRC64:

```

```

Query Match      6 0%: Score 103; DB 1; Length 650;
Best Local Similarity 18.8%: Pred No. 3.8;
Matches 68; Conservative 57; Mismatches 124; Indels 112; Gaps 13;

OY 43 ELARDPSRLDLYVDVYSGAGVIYKPEDMYVDLGINNMSVLLTFSARLQAYVKNV 102
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 61 QALNPINIVFPAKRRIGRRFSDA-----VOEDMKL-----MPFKVILPPGDK----- 104
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

RESULT	12
ITH1_MESAU	
ID	ITH1_MESAU
	STANDARD;
	PRT;
	914 AA

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

-----

CC EMBL; D89285; BAA13938.1; .

DR InterPro; IPR002035; vwfA.

DR Pfam; PF00092; vwa; 1.

DR SMART; SM00327; vWA; 1.

DR PROSITE; PS50234; vwfA; 1.

DR Serine protease inhibitor; Repeat; Signal; Multigene family;

KW Glycoprotein.

FT SIGNAL 1 30

FT CHAIN 31 675

FT PROPER 676 914

FT DOMAIN 293 453

FT CARBOHYD 288 288

FT CARBOHYD 291 291

FT CARBOHYD 591 591

FT BINDING 675 675

FT SEQUENCE 914 AA; 101785 MW; B693FE56FE89E93 CRC64;

-----

Query Match 6.0%; Score 102.5; DB 1; Length 914;

Best Local Similarity 19.1%; Pred. No. 6.5; Mismatches 119; Indels 101; Gaps 15;

Matches 66; Conservative 59; Mismatches 119; Indels 101; Gaps 15;

QY 16 VLEAETDGLAEGSKRAPEGLVLDFAELARDPSSTRDLITNVVDYVYSGASGIKPED- 74

DB 296 LVFVLDIGSGMGGQKVKOTKEALKIL-----GDVKPDS 330

QY 75 ---MYVDLGINNWSVLLTPS--ARLQA---YKNSVPAVAYKSESKRYAGDTLIGRVYL 126

DB 331 FDLVLFGRVQVQSMKSLVPATQANLQAODVFRFRSLAGAT-----NLNGLLIGIEIL 384

QY 127 -----PFSYQSSAMIMPRFKIPFYSGESGQFLGKLDINIKTKREIKYSV-----Y 174

DB 385 NKAQGSHEELSPASIL-----IMLTDEP-----TEETDRSQILKNVRNATRGREFLY 434

QY 175 SLGYEIDLEVLFEEDNGMEYAVSMGTLKFKGADLIWGN-----PVIYINISSRIITD 227

DB 435 NIGFCHDIDPFLFLEMSENLS-----GMAQGIYEDHDATQGLQGFQNVANPRLITD 485

QY 228 DVPNYP---LAASKMKRFKA-----RVSKSHSSKVKYKFLFYVKDLRVLYDKLSV 273

DB 486 VELQYQDSVSLTLQHRKRYQYDGESEIYVAGRIADHKISTEK-----ADVARGEROEF 539

QY 274 SIDSIDISESVFKYVETSG---TESLRKLKAHEPTKRVYLKREKI 315

DB 540 KATCYLDEEEMKILLRERGHMLENHERIMVLYTLQELLARKMKM 584

-----

RESULT 13

CGAA\_CLOBI STANDARD; PRT; 613 AA.

AC 045882;

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PESTICIDAL CRYSTAL-LIKE PROTEIN CRY16A (INSECTICIDAL TOXIN

DE CRYXVIA(A)) (CBM7 MOSQUITOCIDAL TOXIN).

GN CRY16A OR CRYXVIA(A) OR CBM71.

OS Clostridium bifermentans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

CC Clostridium.

OK NCBI\_Taxid=1490;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15 AND 374-388.

RC STRAIN=CH18 / SUBSP. MALAYSIA;

RX MEDLINE=96236023; PubMed=8655486;

RA Barloy F., Delecluse A., Nicolas L., Legacé M.-M.;

RT "Cloning and expression of the first anaerobic toxin gene from Clostridium bifermentans subsp. malaysia, encoding a new mosquitocidal

```

RA  Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA  Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Brodovsky M.,
RA  Kleink H.-P., Friaser C.M., Smith H.O., Wiese C.R., Ventor J.C.;
RA  "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA  jannaschii."
RL  Science 273:1058-1073(1996).
CC  -I- SIMILARITY: TO A.FULGIDUS AF0817.
CC  -----
CC  CC  THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on less
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/).
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: U67596; AAB9574.1; -.
DR  TIGR: MJ1554. -.
KW  Hypothetical protein; Signal; Complete proteome.
FT  SIGNAL 1 20 POTENTIAL.
FT  CHAIN 21 622 HYPOTHETICAL PROTEIN MJ1554.
SQ  SEQUENCE 622 AA; 72282 MW; 86CAA62F3854D73F CRC64;
Query Match 5.98; Score 101.5; DB 1; Length 622;
Best Local Similarity 20.98; Pred. No. 4.6;
Matches 77; Conservative 74; Mismatches 149; Indels 69; Gaps 20;
QY 1 MKRRKSKILFLLSTLV-LFAOETDCLAGSKRAEKGELVLDPAELARPPSSRLDLTVNV 59
DB 1 MKIKAVAFLLSLMTISLFS-----GCVENKPIKESG--NDFKLIPVNSKSNFEFKNTV 54
QY 60 D-----VYSGASGIVKPEEDVVDLGINNWSVLTLPNSARLQAYVNSVYAPAVKSEKR 114
DB 55 ENSIGNIYIVGHSAVSAREVOITVTKSSN---VETSEPEKFSKTVNQVKVDADILK 110
QY 115 YAGDTILVRLVLFSPYSSGSSAMIPPKPIRPS-----GESGNQPLGKLIDNKTKE 168
DB 111 TNGNIT-----AFSONKIKYLIKPLP-PRYAKIIRKINISEGQYLYLTN--NLTIVISM 158
QY 169 IKVSYISLGEYEDLDELFEE-DWNGKEVAYSM---GTLKF--KGWAD-LWISN----- 213
DB 159 NKITSYVNSNBPMPKTIQMDDLNG--SYVDSRLYNCTILVYRKNSIDCPDIWNNKIGYD 217
QY 214 PNYIENISSRIITKDQVNPVPLASSKMRKAFKFS-----KSHSKSKVKNFIYVNDL 264
DB 218 KYIPELRLPIYSMDPDTYIIT--SRINIKSGCQVENSIAIVGNKYTTIVSKNNKLFAVNL 275
QY 285 RVLVDKLSVSDSIDSSVEFKVYETSTGTESTIRLKAKAHEFEKRVYLKLRKISIAEGSFON 324
DB 276 KINEKKLML-----NFLNESADKYPFTEVADKIKRVIENEDFGDNAKF--VEITE-TTER 327
QY 325 EVEKIESER 333
DB 328 YLSSLPSEK 336
RESULT 15
HS7C_TRYBB
ID HS7C_TRYBB STANDARD. PRT; 676 AA.
AC P20030;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE HEAT SHOCK COGNATE HSP70 PROTEIN.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxId=5702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90377290; PubMed=239817;
RA Lee M.G.-S., Polvere R.I., van der Ploeg L.H.T.;
RA "Evidence for segmental gene conversion between a cognate hsp 70 gene

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2002, 01:58:06 ; Search time 2486.21 Seconds

(Without alignments)  
139.345 Million cell updates/sec

Title: US-09-004-395-3

Sequence: 1 atgaaaggaagctaaagt 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vl:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.om:\*  
20: em.or:\*  
21: em.ov:\*  
22: em.pat:\*  
23: em.ph:\*  
24: em.pl:\*  
25: em.ro:\*  
26: em.sts:\*  
27: em.sy:\*  
28: em.un:\*  
29: em.vl:\*  
30: em.htgo.hum:\*  
31: em.htgo.inv:\*  
32: em.htgo.rod:\*  
33: em.htg.hum:\*  
34: em.htg.inv:\*  
35: em.htg.rod:\*  
36: em.htg.other:\*



Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	1655	1	BB062900
2	21	100.0	13271	1	AE001168
3	20	95.2	3885	8	LEPUR
4	19.4	92.4	183289	2	AC068117
5	19.4	92.4	198207	2	AC068539
6	18.4	87.6	33045	3	CET25C12
7	18.4	87.6	62555	2	AC091103
8	18.4	87.6	139064	2	AC073440
9	18.4	87.6	148142	9	AC011421
10	18.4	87.6	156048	2	AC091111
11	18.4	87.6	159012	2	AC023145
12	18.4	87.6	161624	2	AP001493
13	18.4	87.6	164652	2	AC026923
14	18.4	87.6	170270	2	AP001548
15	18.4	87.6	170413	2	AC090405
16	18.4	87.6	175178	2	AC011783
17	18.4	87.6	175850	2	AP001399
18	18.4	87.6	175924	2	AC092883
19	18.4	87.6	176421	2	AC025157
20	18.4	87.6	194487	2	AP002896
21	18	85.7	5954	10	AF247181
22	18	85.7	27697	2	AC017568
23	18	85.7	62238	2	AC012390
24	18	85.7	73360	9	HS06083
25	18	85.7	112621	9	AC004891
26	18	85.7	159570	3	AC010667
27	18	85.7	169572	2	AC023832
28	18	85.7	199891	9	CNS00080
29	18	85.7	320565	3	AE003844
30	17.8	84.8	2794	5	XL019974
31	17.8	84.8	10255	1	AC015201
32	17.8	84.8	13164	1	AE006153
33	17.8	84.8	14733	2	AC014738
34	17.8	84.8	24137	9	AL157368
35	17.8	84.8	29034	9	HS243947
36	17.8	84.8	40397	8	SC99359
37	17.8	84.8	82289	8	ATAC009895
38	17.8	84.8	97912	9	HSJ753D10
39	17.8	84.8	111641	9	HS232N11
40	17.8	84.8	112755	10	AC083887
41	17.8	84.8	113253	2	AC092356
42	17.8	84.8	115884	2	AC017581
43	17.8	84.8	116518	2	AC067732
44	17.8	84.8	117453	3	AE003670
45	17.8	84.8	120314	9	AL133338

## ALIGNMENTS

RESULT 1	BB062900	1655 bp	DNA	BCT	15-JAN-1997
LOCUS	BB062900				
DEFINITION	Borrelia burgdorferi flagellar filament outsheath protein (flaA) gene, complete cds, and chemotaxis histidine kinase (cheA) gene, partial cds.				
ACCESSION	U62900				
VERSION	U62900.1	GI:1575445			
KEYWORDS	Lyme disease spirochete.				
SOURCE	Borrelia burgdorferi				
ORGANISM	Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.				
REFERENCE	1 (bases 1 to 1655)				
AUTHORS	Ge, Y. and Charon, N.W.				
TITLE	An unexpected flaA homolog is present and expressed in Borrelia burgdorferi				
JOURNAL	J. Bacteriol. 179 (2), 552-556 (1997)				
MEDLINE	97144545				
REFERENCE	2 (bases 1 to 1655)				

AUTHORS Ge, Y.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-1996) Yigong Ge, West Virginia University,  
Microbiology, HSCN, Morgantown, WV 26506, USA  
FEATURES  
source  
1. 1655  
/organism="Borrelia burgdorferi"  
/strain="212"  
/db\_xref="taxon:139"  
/clone="pm1 and pm2"  
1. 395  
/function="unknown"  
/note="oriA"  
/codon\_start=3  
/transl\_table=1  
/protein\_id="AAC4769.1"  
/db\_xref="GI:1575446"  
/translation="DNLFKRVYFPHSSKGFVANIRYLRDQNLKDNLDLVKDFLLG  
SNEGFSEFLLSDSNFLSLKNGVYVNLREFYDSFNNGDYNESFDVKVNLFPAMSL  
IKTRFNYPGKIKRIIIVEGILKEOS"  
473. 1498  
/gene="flaA"  
473. 1498  
/gene="flaA"  
/function="putative flagellar filament outsheath protein"  
/codon\_start=1  
/transl\_table=1  
/product="flaA protein"  
/protein\_id="AAC44770.1"  
/db\_xref="GI:1575447"  
/translation="WKRKASILFFELSTVLEPAETDGLAEGSKRAEPGELVDFAE  
LRDPSSTRDLTNVYDYVYSGASGIVKPEDVVDGIMNWSVLLTPSARLQAYVNSV  
VAPAVKSESKRYAGDTLLGVLEPYSOSAMIMPEFKIPEYSGESGNGFLGKI  
DNITMKIEIKVSVYSLAGEIDLEVLFEEDMNGMEVYASMGTLKFGMADLIMSNPYIP  
NISRIITKDDVPNPLASSKMRKAFRYSKSHSSKSKNFITYVDLRLVLYKLSYSD  
SDIDSESVFKYETSGTSLRKLKAHETFKVLRKREKISTAGSFQNFVEKISEKP  
EESSEPKN"  
gene  
1566. 1655  
/gene="cheA"  
1566. >1655  
/gene="cheA"  
/function="chemotaxis histidine kinase"  
/codon\_start=1  
/transl\_table=1  
/product="CheA protein"  
/protein\_id="AAC44771.1"  
/db\_xref="GI:1575448"  
/translation="WEILDLENEELLGVFEPAQNLVDILEENT"  
BASE COUNT 560 a 182 c 308 g 605 t  
ORIGIN  
Query Match 100.0%; Score 21; DB 1; Length 1655;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 acgaaaggaagcctaaagt 21  
|||||  
Db 473 ATGAAAGGAAGCAAGCTAAAGT 493  
RESULT 2  
AE001168 13271 bp DNA BCT 15-DEC-1997  
LOCUS Borrelia burgdorferi (section 54 of 70) of the complete genome.  
ACCESSION AE001168 AE000783  
VERSION AE001168.1 GI:2688598  
KEYWORDS  
SOURCE Lyme disease spirochete.  
ORGANISM Borrelia burgdorferi  
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
burgdorferi group.  
REFERENCE 1 (bases 1 to 13271)

AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.,  
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
Gwinn,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D., Richardson,D.,  
Peterson,J., Kerlavage,A.R., Quackenbush,J., Salzberg,S.,  
Hanson,M., van Vugt,R., Palmer,N., Adams,M.D., Gocayne,J.,  
Venter,J.C., et al.  
TITLE Genomic sequence of a Lyme disease spirochaete, Borrelia  
burgdorferi  
JOURNAL Science 290 (6660), 580-586 (1997)  
MEDLINE 98065943  
REFERENCE 2 (bases 1 to 13271)  
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,  
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
Gwinn,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D.,  
Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,  
Salzberg,S., Hanson,M., van Vugt,R., Palmer,N., Adams,M.D.,  
Gocayne,J.D., Weidman,J., Utterback,T., Wattney,L., McDonald,L.,  
Artach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,  
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.  
TITLE Direct Submission  
JOURNAL Submitted (12-DEC-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
source  
1. 13271  
/organism="Borrelia burgdorferi"  
/db\_xref="taxon:139"  
gene  
66. 1100  
/gene="BB0668"  
66. 1100  
/gene="BB0668"  
/note="similar to GP:1575447 percent identity: 98.83;  
identified by sequence similarity: putative"  
/codon\_start=1  
/transl\_table=1  
/product="flagellar filament outer layer protein (flaA)"  
/protein\_id="AAC67025.1"  
/db\_xref="GI:2688608"  
/translation="MYMKRRAKSLFFELSTVLEPAETDGLAEGSKRAEPGELVDF  
AEIARDPSTRDLTNVYDYVYSGASGIVKPEDVVDGIMNWSVLLTPSARLQAYVNSV  
NSVAPAVKSESKRYAGDTLLGVLEPYSOSAMIMPEFKIPEYSGESGNGFLGKI  
GLDINKIKRKIEIKVSVYSLAGEIDLEVLFEEDMNGMEVYASMGTLKFGMADLIMSNPN  
YIPNISRIITKDDVPNPLASSKMRKAFRYSKSHSSKSKNFITYVDLRLVLYKLSY  
SDIDSESVFKYETSGTSLRKLKAHETFKVLRKREKISTAGSFQNFVEKISEKS  
EKPESSPKN"  
gene  
1168. 3762  
/gene="BB0669"  
1168. 3762  
/gene="BB0669"  
/note="similar to PID:1113815 SP:Q44737 GB:AE000783  
percent identity: 99.31; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=1  
/product="chemotaxis histidine kinase (cheA-2)"  
/protein\_id="AAC67024.1"  
/db\_xref="GI:2688607"  
/translation="WEILDLENEELLGVFEPAQNLVDILEENISLEDDPNNSDTID  
EIPRAHRLKGSASISDMMELSDPHTIYEDVDAIRDKGVNINDVLLSLSDVVK  
EMALRIDGKYINDISDLKSKIKQFLYIDQTFIKRDSGINSIKNNRGLSDSEIR  
EGLGIGQVLRISVFNSSNSSEVENSGLAKFINILKNGSVLHTIPKEQIIDEKFLK  
RVDYLLYSDIEGVKSLDSNLIESYLVDFENKBEIKRLADEIKRVDLIDENFVN  
DNPDFTEDEIDLEVENOKLFYRDLFVDNDPMATISGLQMLAKSLKIRFSIP  
DSSELADKPEDFYVYILISNTSEESIAKKINLPDVYSFEIKVNLSELSKRLKED  
DEAPREKNNKIKNSPTSVNLIRIDSKRIDVILNIVSAVYSKSYNGINSEMTLFP  
NPNFYIDQESFORNRLDKIVRDAKLTIDELESHINSLSFKMAKAKDISELR  
NSFFRLQNFMTSGRLSRITTDLHESYKLRMLPISNIFSRFVAVDLSKRLKIV  
NLKMGDETELDGKVIDVLPRLKLCVNSNDHDELEVYEVKRGKSAAGTIIIRAKN  
EGNVYSIEIEDGIGIDPKYIRRLKIEGTIKEDAIYSDPELINLIFAPGFTAVQV  
DLSGGVGLDVVKKSIERLNGTILVESIGTIFPKIKLPLTVIIOGLVKSGETY  
VIPLNNVETFRIRIETHDKILENHEYNLDEDEVYISVLRKLNFTNPDDSLIEKELI  
VNTSNMKAIIVDSIIGEDPFVYKPIKDKFSSSAGIYGATTLGNGKVVLLIIVFKFL  
DLQDKTK"  
3769. 5169  
gene

gene  
CDS  
/gene="BB0670"  
3769..5169  
/note="similar to GP:1881576 percent identity: 98.93;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="purine-binding chemotaxis protein (chew-3)"  
/protein\_id="AAC67023.1"  
/db\_xref="GI:2688606"  
/translation="MQIKETIYFGRILDDKNSKRLTDFEKFVSEFELGSDHYLVDM  
QVKEIRKSNSEYVFNARKYVAGLDNLRGEIIPIDLRIMPLEFNKDIEDIMVAKN  
EDLIGVLYDKINNVESIDSLIDPHVPLSODSLINIKGVADNEKYLILANFKI  
FNVEGEKILKPGONFVERSDPVDNDLLENSKNDSPKSSKDNNSNTLN  
TAPVLENIKKRLKLYSRNASLVNDFLEKVCYKFMVDTDLPCDSFLNERYSKSSG  
NMGADCLSEERKNEIYESRSLSMNNLSIFNVLEIGCGSGKETMALANLSEYKPF  
KLTAIDNLSKVERSLVSESEIGSEIYSRNSFQSPGVYKFKSELSNVLEYSS  
DALFSDLPDLNGLVFLKDVLCFLDSKDOIILNITASKAIKGLALVLGDNELKNDV  
FIKESKTYFNLYRKV"  
5181..5666  
/gene="BB0671"  
5181..5666  
/note="similar to GP:1408273 percent identity: 98.76;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis operon protein (chex)"  
/protein\_id="AAC67022.1"  
/db\_xref="GI:2688605"  
/translation="MRIDYIEPFLDAASSVLKMDLVNENIEMKQGLKSIKNIKGV  
VIVGLAGSEGIITIDMDIETFLPAASKINFEEDFDEDEKEMVAALTEKGNITIA  
GNEVTLAKGKGFVDITPPAFIYGENMKISNGSEALIVPFSLPDKIIEVMAIRER  
V"  
5706..6146  
/gene="BB0672"  
5706..6146  
/note="similar to GP:1408274 percent identity: 97.95;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis response regulator (chex-3)"  
/protein\_id="AAC67021.1"  
/db\_xref="GI:2688604"  
/translation="MIQTTIADSSSRFGINYPNVLIVDSVFTVKOLTOI  
FTSEGFNIIDPADGEAVIKYKNHPNIDITLDTMKMGITCLSNIMEFDKNAR  
VIMISALGKQELVCLIKGAKFTYKPLDRAKVLQRVMSVFEK"  
complement(6187..6702)  
/gene="BB0673"  
complement(6187..6702)  
/note="similar to GP:1408275 percent identity: 97.56;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC67020.1"  
/db\_xref="GI:2688603"  
/translation="MKIQIILMLALDLPLNARLLDISIEKRADEIKKSSVNIIL  
KEEYVNTPTSEIKENIKYKLTSEHFWKSLNKNKNSKKNKVKYIOSELIDK  
PLKVKIFIKINNGISLITTKGPKLELILYENNAEPLKIRNLNITTYELKKNIDK  
ISNEMIFPERE"  
complement(6699..7745)  
/gene="BB0674"  
complement(6699..7745)  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0674"  
/protein\_id="AAC67027.1"

gene  
CDS  
/db\_xref="GI:2688610"  
/translation="MWKRSIFLKAIIIFSEFLIEELSIILFLPYKIRPALIFLGF  
LDDYPIPELKKIRKAYLSORLEIVVRNNLEDDITHCLIPAFSSVOLKNIVANE  
TILNPIMLSFKRLRLRLRNDLIIEIYNSKEKMLIILARFTSMALLPPTFFI  
IISSSKIVSIEPEKQENIKINISINERAVIKETPFLIIKEKDDIYKSDIEIV  
YVSPSEYVIEEMKTFYIDKYLORSDSILGFLTLPLASFLIPLMNFYKFKKASFL  
NPILMTKILQDPLEYRKIQIPFTLSEKVEYELAKSFNMLLKEKLNKRSKIPLEI  
EKVKIRINKNOEK"  
complement(7739..8602)  
/gene="BB0675"  
complement(7739..8602)  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0675"  
/protein\_id="AAC67026.1"  
/db\_xref="GI:2688609"  
/translation="MNFKPYFPLVLYLFNHSIKSLASIEKEELIYLNKKNKRTIA  
TIRKNDRESEIKDLOIYVAKDKDIMTPPYVLSIEADINIKYPLFSKELTKNDLNI  
FNFVKVNSSLRKEEFYFNNTISNGYTFYINKLFEKNSYTYILIQENKALSSDI  
KNYIKILLKLVKIVIKYCEKEGIELTKNIESTSKAISDTDFLDEKTAKLIISEFFK  
YETLOTMSPISTLIIAIFSARARPKYKNPVKGIIGYDESMFSIKQSGSREYDSRIIK  
ELSEIAKVNKW"  
complement(8599..9261)  
/gene="BB0676"  
complement(8599..9261)  
/note="similar to GB:142023.SP:P44755.PID:1004013  
PID:1222502.PID:1204815 percent identity: 25.58;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoglycolate phosphatase (gph)"  
/protein\_id="AAC67019.1"  
/db\_xref="GI:2688602"  
/translation="MKIKACIPDMOSTLVNSINDIAPSNKSNALSNIGYSKIELSKENA  
LVGRGNKRFVIDTLKLSLEHDPNLOEKLKFEVYENKNSLSFOKTPENIKPLET  
MKNLNPICIGLSKNHEILINLVKNIGNITLFEFIEGYSKNEPFPKDPNALMILET  
NAQKEIIAYIGSDVDMLTALNMGFPGISVWGFSVOELKESGAKHIIHNPLELIDL  
IK"  
9504..11114  
/gene="BB0677"

Query Match 100.0%; Score 21; DB 1; Length 13271;  
Best local Similarity 100.0%; Pred. No. 47;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	atgaaagagaagctaaagt 21
Db	75	ATGAAAGGAAGCTAAAGT 95

RESULT 3

LEPUR	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
LEPUR	3885 bp	DNA	PLN	03-NOV-1994	esculentum polyubiquitin repeats.	Esculentum polyubiquitin gene from tomato.	Esculentum polyubiquitin gene; ubiquitin.

REFERENCE 1 (bases 1 to 3885)  
AUTHORS Rollink, J.K. and Piltzner, U.M.  
TITLE Structure of a heptaubiquitin gene from tomato  
JOURNAL Plant Physiol. 104 (1), 299-300 (1994)  
MEDLINE 94159799  
REFERENCE 2 (bases 1 to 3885)





\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1287: contig of 1287 bp in length  
1288 1387: gap of 100 bp  
1388 3151: contig of 1764 bp in length  
3152 3251: gap of 100 bp  
3252 4709: contig of 1458 bp in length  
4710 4809: gap of 100 bp  
4810 6242: contig of 1433 bp in length  
6243 6342: gap of 100 bp  
6343 8667: contig of 2325 bp in length  
8668 8767: gap of 100 bp  
8768 10535: contig of 1768 bp in length  
10536 10635: gap of 100 bp  
10636 12007: contig of 1372 bp in length  
12008 12107: gap of 100 bp  
12108 14409: contig of 2302 bp in length  
14410 14509: gap of 100 bp  
14510 16536: contig of 2027 bp in length  
16537 16636: gap of 100 bp  
16637 19520: contig of 2884 bp in length  
19521 19620: gap of 100 bp  
19621 22669: contig of 3049 bp in length  
22670 22769: gap of 100 bp  
22770 26259: contig of 3490 bp in length  
26260 26359: gap of 100 bp  
26360 29359: contig of 3000 bp in length  
29360 29459: gap of 100 bp  
29460 31820: contig of 2361 bp in length  
31821 31920: gap of 100 bp  
31921 34836: contig of 2916 bp in length  
34837 34936: gap of 100 bp  
34937 38183: contig of 3247 bp in length  
38184 38283: gap of 100 bp  
38284 40743: contig of 2460 bp in length  
40744 40843: gap of 100 bp  
40844 46140: contig of 5297 bp in length  
46141 46240: gap of 100 bp  
46241 52339: contig of 6099 bp in length  
52340 52439: gap of 100 bp  
52440 57689: contig of 5250 bp in length  
57690 57789: gap of 100 bp  
57790 65948: contig of 8159 bp in length  
65949 66048: gap of 100 bp  
66049 74025: contig of 7977 bp in length  
74026 74125: gap of 100 bp  
74126 80421: contig of 6296 bp in length  
80422 80521: gap of 100 bp  
80522 89984: contig of 9463 bp in length  
89985 90084: gap of 100 bp  
90085 99029: contig of 8945 bp in length  
99030 99129: gap of 100 bp  
99130 109926: contig of 10797 bp in length  
109927 110026: gap of 100 bp  
110027 120095: contig of 10069 bp in length  
120096 120195: gap of 100 bp  
120196 131804: contig of 11609 bp in length  
131805 131904: gap of 100 bp  
131905 142878: contig of 10974 bp in length  
142879 142978: gap of 100 bp  
142979 153656: contig of 10678 bp in length  
153657 153756: gap of 100 bp  
153757 169035: contig of 15279 bp in length  
169036 169135: gap of 100 bp  
169136 183289: contig of 14154 bp in length.

Location/Qualifiers  
1.183289  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"

FEATURES  
source

misc\_feature /clone="RP11-353K11"  
1.1287 /clone\_11b="RP11 Human Male BAC"  
misc\_feature /note="assembly\_fragment"  
1388.3151  
misc\_feature /note="assembly\_fragment"  
3252.4709  
misc\_feature /note="assembly\_fragment"  
4810.6242  
misc\_feature /note="assembly\_fragment"  
6343.8667  
misc\_feature /note="assembly\_fragment"  
8768.10535  
misc\_feature /note="assembly\_fragment"  
10636.12007  
misc\_feature /note="assembly\_fragment"  
12108.14409  
misc\_feature /note="assembly\_fragment"  
14510.16536  
misc\_feature /note="assembly\_fragment"  
16637.19520  
misc\_feature /note="assembly\_fragment"  
19621.22669  
misc\_feature /note="assembly\_fragment"  
22770.26259  
misc\_feature /note="assembly\_fragment"  
26360.29359  
misc\_feature /note="assembly\_fragment"  
29460.31820  
misc\_feature /note="assembly\_fragment"  
31921.34836  
misc\_feature /note="assembly\_fragment"  
34937.38183  
misc\_feature /note="assembly\_fragment"  
38284.40743  
misc\_feature /note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
40844.46140  
misc\_feature /note="assembly\_fragment"  
46241.52339  
misc\_feature /note="assembly\_fragment"  
52440.57689  
misc\_feature /note="assembly\_fragment"  
57790.65948  
misc\_feature /note="assembly\_fragment"  
66049.74025  
misc\_feature /note="assembly\_fragment"  
74126.80421  
misc\_feature /note="assembly\_fragment"  
80522.89984  
misc\_feature /note="assembly\_fragment"  
90085.99029  
misc\_feature /note="assembly\_fragment"  
99130.109926  
misc\_feature /note="assembly\_fragment"  
110027.120095  
misc\_feature /note="assembly\_fragment"  
120196.131804  
misc\_feature /note="assembly\_fragment"

Query Match 92.4% Score 19.4: DB 2: Length 183289;  
Best Local Similarity 95.2% Pred.No.2.4e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgaagaagaaagctaaagt 21  
|||||  
Db 52810 ATGAAAGAAACTATAAAGT 52790

RESULT 5  
AC068539 AC068539 198207 bp DNA HTG 12-AUG-2000

DEFINITION Homo sapiens chromosome 2 clone RP11-145M22, WORKING DRAFT  
SEQUENCE, 13 unordered pieces.  
ACCESSION AC068539  
VERSION AC068539.5 GI:9719795  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE 1 (bases 1 to 198207)  
JOURNAL Waterston,R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 198207)  
Waterston,R.H.  
REFERENCE Waterston,R.H.  
AUTHORS Direct Submission  
TITLE Submitted (03-MAY-2000) Genome Sequencing Center, Washington  
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Aug 7, 2000 this sequence version replaced gi:8571789.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H.NH0145M22  
----- Summary Statistics -----  
Sequencing vector: M13, 100%  
Sequencing vector: Plasmid; 0%  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 190362 bases at least Q40  
Consensus quality: 194318 bases at least Q30  
Insert size: 20000; agarose-fp  
Insert size: 197007; sum-of-ctrls  
Quality coverage: 4.96 in Q20 bases; agarose-fp  
Quality coverage: 5.07 in Q20 bases; sum-of-ctrls

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
\* 2651: contig of 2651 bp in length  
\* 2652: gap of unknown length  
\* 2752: contig of 2781 bp in length  
\* 5532: gap of unknown length  
\* 5533: contig of 5703 bp in length  
\* 11335: gap of unknown length  
\* 11336: contig of 6705 bp in length  
\* 11436: gap of unknown length  
\* 18141: gap of unknown length  
\* 18241: gap of unknown length  
\* 25599: contig of 7359 bp in length  
\* 25600: gap of unknown length  
\* 25700: gap of unknown length  
\* 36129: contig of 10429 bp in length  
\* 36228: gap of unknown length  
\* 36229: contig of 12825 bp in length  
\* 49053: gap of unknown length  
\* 49054: gap of unknown length  
\* 49154: contig of 12234 bp in length  
\* 61388: gap of unknown length  
\* 61488: contig of 12836 bp in length  
\* 74323: gap of unknown length  
\* 74324: contig of 14470 bp in length  
\* 88893: gap of unknown length  
\* 88894: contig of 19852 bp in length  
\* 108945: gap of unknown length  
\* 108946: contig of 34470 bp in length  
\* 143416: gap of unknown length

FEATURES \* 143516 198207: contig of 34692 bp in length.  
source  
1..198207  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-145M22"  
1..2651  
misc\_feature  
/note="assembly\_name:Contig19"  
2752..5532  
/note="assembly\_name:Contig20"  
5533..11335  
/note="assembly\_name:Contig21"  
11336..18140  
/note="assembly\_name:Contig22"  
18241..25599  
/note="assembly\_name:Contig23"  
25700..36128  
/note="assembly\_name:Contig24"  
36229..49053  
/note="assembly\_name:Contig25"  
49154..61387  
/note="assembly\_name:Contig26  
clone\_end:SP6  
vector\_side:right"  
61488..74323  
/note="assembly\_name:Contig27"  
74424..88893  
/note="assembly\_name:Contig28"  
88994..108845  
/note="assembly\_name:Contig29  
clone\_end:T7  
vector\_side:left"  
108946..143415  
/note="assembly\_name:Contig30"  
143516..198207  
/note="assembly\_name:Contig31"  
BASE COUNT 51443 a 47183 c 48061 g 50315 t 1205 others  
ORIGIN

Query Match 92.4%; Score 19.4; DB 2; Length 198207;  
Best Local Similarity 95.2%; Pred. No. 2.4e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 atgaaggaagcraaagt 21  
|||||  
Db 103013 ATGAAAGGAAACTTAAAGT 103033

RESULT 6  
LOCUS CEM25C12 33045 bp DNA INV 20-JUN-2001  
DEFINITION Caenorhabditis elegans cosmid T25C12, complete sequence.  
ACCESSION Z66566  
VERSION Z66566.1 GI:1051339  
KEYWORDS HTG; lin-14.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;  
AUTHORS Rhabditiida; Rhabditiidae; Peloderiinae; Caenorhabditis.  
TITLE 1 (bases 1 to 33045)  
JOURNAL none.  
MEDLINE Genome sequence of the nematode C. elegans: a platform for  
REMARK Investigating biology. The C. elegans Sequencing Consortium  
AUTHORS Science 282 (5396), 2012-2018 (1998)  
TITLE The C. elegans Sequencing Consortium.  
JOURNAL 2 (bases 1 to 33045)  
Kershaw,J.K.  
Direct Submission  
Submitted (04-NOV-1995) Nematode Sequencing Project, Sanger Centre,  
Hinxton, Cambridge CB10 1RO, England and Department of Genetics,



comes from this gene  
 cDNA EST yk22a7.3 comes from this gene; cDNA EST yk26b2.3  
 comes from this gene  
 cDNA EST yk38a1.3 comes from this gene; cDNA EST yk42e7.3  
 comes from this gene  
 cDNA EST yk13d10.5 comes from this gene; cDNA EST yk2h10.5  
 comes from this gene  
 cDNA EST yk4b3.5 comes from this gene; cDNA EST yk5f7.5  
 comes from this gene  
 cDNA EST yk7a3.5 comes from this gene; cDNA EST yk13c12.5  
 comes from this gene  
 cDNA EST yk10b4.5 comes from this gene; cDNA EST yk23b5.5  
 comes from this gene  
 cDNA EST yk26b2.5 comes from this gene; cDNA EST yk38a1.5  
 comes from this gene  
 cDNA EST yk22a7.5 comes from this gene; cDNA EST  
 yk38b10.3 comes from this gene  
 cDNA EST yk38b10.5 comes from this gene; cDNA EST  
 yk259f5.3 comes from this gene  
 cDNA EST yk259f5.5 comes from this gene; cDNA EST  
 yk295a1.3 comes from this gene  
 cDNA EST yk295a1.5 comes from this gene; cDNA EST  
 yk294g11.3 comes from this gene  
 cDNA EST yk294g11.5 comes from this gene; cDNA EST  
 yk285b2.3 comes from this gene  
 cDNA EST yk285b2.5 comes from this gene; cDNA EST  
 yk276d2.3 comes from this gene  
 cDNA EST yk276d2.5 comes from this gene; cDNA EST  
 yk243d6.3 comes from this gene  
 cDNA EST yk243d6.5 comes from this gene; cDNA EST  
 yk239d3.3 comes from this gene  
 cDNA EST yk239d3.5 comes from this gene; cDNA EST  
 yk226g10.3 comes from this gene

Query Match 87.6%; Score 18.4; DB 3; Length 33045;  
 Best Local Similarity 95.0%; Pred. No. 6,1e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 atgaagaaggaagctaaag 20  
 |||||||||  
 DB 1344 ATGAAAGCAAACTAAG 1325

RESULT 7  
 AC091103 62555 bp DNA HTG 29-MAR-2001  
 LOCUS Homo sapiens chromosome 18 clone RP11-749G1 map 18, LOW-PASS  
 DEFINITION  
 AC091103  
 AC091103  
 AC091103.1 GI:13488002  
 VERSION HTG; HTGS\_PHASE0.  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 62555)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 18, clone RP11-749G1  
 Unpublished  
 2 (bases 1 to 62555)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
 Bartha,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,  
 Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,  
 Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,  
 Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,  
 Iliev,I., Johnson,R., Jones,C., Karatas,A., Larocque,K.,  
 Iamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
 Maclean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,  
 McEwan,P., McKernan,K., McPheters,R., Meldrum,J., Menes,L.,  
 Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nordu,C.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,

# TITLE JOURNAL COMMENT

Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
 Roselli,M., Roy,A., Santos,R., Schaner,S., Schnuppback,R., Seaman,S.,  
 Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,  
 Theodore,J., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (29-MAR-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence.submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: 749\_G1  
 Center clone name: 749\_G1  
 -----  
 \* NOTE: This record contains 70 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.  
 \* 1 795: contig of 795 bp in length  
 \* 796 895: gap of 100 bp  
 \* 896 1636: contig of 741 bp in length  
 \* 1637 1736: gap of 100 bp  
 \* 1737 2537: contig of 801 bp in length  
 \* 2538 2637: gap of 100 bp  
 \* 2638 3408: contig of 771 bp in length  
 \* 3409 3508: gap of 100 bp  
 \* 3509 4313: contig of 805 bp in length  
 \* 4314 4413: gap of 100 bp  
 \* 4414 5220: contig of 807 bp in length  
 \* 5221 5320: gap of 100 bp  
 \* 5321 6125: contig of 805 bp in length  
 \* 6126 6225: gap of 100 bp  
 \* 6226 6993: contig of 768 bp in length  
 \* 6994 7093: gap of 100 bp  
 \* 7094 7891: contig of 798 bp in length  
 \* 7892 7991: gap of 100 bp  
 \* 7992 8783: contig of 792 bp in length  
 \* 8784 8883: gap of 100 bp  
 \* 8884 9674: contig of 791 bp in length  
 \* 9675 9774: gap of 100 bp  
 \* 9775 10576: contig of 802 bp in length  
 \* 10577 10676: gap of 100 bp  
 \* 10677 11486: contig of 810 bp in length  
 \* 11487 11586: gap of 100 bp  
 \* 11587 12384: contig of 798 bp in length  
 \* 12385 12484: gap of 100 bp  
 \* 12485 13265: contig of 781 bp in length  
 \* 13266 13365: gap of 100 bp  
 \* 13366 14161: contig of 796 bp in length  
 \* 14162 14261: gap of 100 bp  
 \* 14262 15053: contig of 792 bp in length  
 \* 15054 15153: gap of 100 bp  
 \* 15154 15955: contig of 802 bp in length  
 \* 15956 16055: gap of 100 bp  
 \* 16056 16864: contig of 809 bp in length  
 \* 16865 16964: gap of 100 bp  
 \* 16965 17773: contig of 809 bp in length

17774 17873: gap of 100 bp  
17874 18705: contig of 832 bp in length  
18706 18805: gap of 100 bp  
18806 18805: contig of 824 bp in length  
18806 18805: gap of 100 bp  
19306 19729: gap of 100 bp  
19730 20538: contig of 809 bp in length  
20539 20638: gap of 100 bp  
20639 21461: contig of 823 bp in length  
21462 21561: gap of 100 bp  
21562 22380: contig of 819 bp in length  
22381 22480: gap of 100 bp  
22481 23278: contig of 798 bp in length  
23279 23378: gap of 100 bp  
23379 24169: contig of 791 bp in length  
24170 24269: gap of 100 bp  
24270 25056: contig of 787 bp in length  
25057 25156: gap of 100 bp  
25157 25971: contig of 815 bp in length  
25972 26071: gap of 100 bp  
26072 26864: contig of 793 bp in length  
26865 26964: gap of 100 bp  
26965 27778: contig of 814 bp in length  
27779 27878: gap of 100 bp  
27879 28661: contig of 783 bp in length  
28662 28761: gap of 100 bp  
28762 29545: contig of 784 bp in length  
29546 29645: gap of 100 bp  
29646 30418: contig of 773 bp in length  
30419 30518: gap of 100 bp  
30519 31296: contig of 778 bp in length  
31297 31396: gap of 100 bp  
31397 32204: contig of 808 bp in length  
32205 32304: gap of 100 bp  
32305 33095: contig of 791 bp in length  
33096 33195: gap of 100 bp  
33196 33982: contig of 787 bp in length  
33983 34082: gap of 100 bp  
34083 34886: contig of 804 bp in length  
34887 34986: gap of 100 bp  
34987 35799: contig of 813 bp in length  
35800 35899: gap of 100 bp  
35900 36689: contig of 790 bp in length  
36690 36789: gap of 100 bp  
36790 37593: contig of 804 bp in length  
37594 37693: gap of 100 bp  
37694 38486: contig of 793 bp in length  
38487 38586: gap of 100 bp  
38587 39386: contig of 800 bp in length  
39387 39486: gap of 100 bp  
39487 40281: contig of 795 bp in length  
40282 40381: gap of 100 bp  
40382 41179: contig of 798 bp in length  
41180 41279: gap of 100 bp  
41280 42094: contig of 815 bp in length  
42095 42194: gap of 100 bp  
42195 42992: contig of 798 bp in length  
42993 43092: gap of 100 bp  
43093 43887: contig of 795 bp in length  
43888 43987: gap of 100 bp  
43988 44752: contig of 765 bp in length  
44753 44852: gap of 100 bp  
44853 45626: contig of 774 bp in length  
45627 45726: gap of 100 bp  
45727 46507: contig of 781 bp in length  
46508 46607: gap of 100 bp  
46608 47403: contig of 796 bp in length  
47404 47503: gap of 100 bp  
47504 48253: contig of 750 bp in length  
48254 48353: gap of 100 bp  
48354 49177: contig of 824 bp in length  
49178 49277: gap of 100 bp  
49278 50094: contig of 817 bp in length  
50095 50194: gap of 100 bp

50195 50959: contig of 765 bp in length  
50960 51059: gap of 100 bp  
51060 51849: contig of 790 bp in length  
51850 51949: gap of 100 bp  
51950 52741: contig of 792 bp in length  
52742 52841: gap of 100 bp  
52842 53635: contig of 794 bp in length  
53636 53735: gap of 100 bp  
53736 54476: contig of 741 bp in length  
54477 54576: gap of 100 bp  
54577 55375: contig of 799 bp in length  
55376 55475: gap of 100 bp  
55476 56258: contig of 783 bp in length  
56259 56358: gap of 100 bp  
56359 57161: contig of 803 bp in length  
57162 57261: gap of 100 bp  
57262 58053: contig of 792 bp in length  
58054 58153: gap of 100 bp  
58154 58959: contig of 806 bp in length  
58960 59059: gap of 100 bp  
59060 59865: contig of 806 bp in length  
59866 59965: gap of 100 bp  
59966 60765: contig of 800 bp in length  
60766 60865: gap of 100 bp  
60866 61660: contig of 795 bp in length  
61661 61760: gap of 100 bp

Query Match 87.6%; Score 18.4; DB 2; Length 62555;  
Best Local Similarity 95.0%; Pred. No. 6.2e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 tgaagaagaaagctaaagt 21  
Db 46462 TGAAAAGCAAGCTAAAGT 46481  
||||||| |||||||||

RESULT 8  
AC073440/C  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC073440 139064 bp DNA HTG 07-SEP-2000  
Homo sapiens chromosome 12 clone RP11-100E14, WORKING DRAFT  
SEQUENCE, 7 unordered pieces.  
AC073440  
AC073440.10 GI:9972211  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 139064)

REFERENCE  
AUTHORS

Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,  
Bodola,B., Bouck,J., Bowie,S., Brooks,A., Bunay,C., Bunac,C.,  
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,  
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,  
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,  
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,  
Guerrero,W., Harris,K., Hernandez,U., Hodgson,A., Hughes,M.,  
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,  
Kelly,S., Kondejowski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,  
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,  
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M.,  
Morris,S., Naeih,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,  
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,  
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,  
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugeng,R.,  
Taber,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,  
Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,  
Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and  
Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 139064)  
Worley,K.C.  
Direct Submission

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE

JOURNAL Submitted (18-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Sep 6, 2000 this sequence version replaced gi:9964656.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HBCF  
Center clone name: RP11-100E14  
----- Summary Statistics  
Sequencing vector: M13; L08821  
Chemistry: Dye-terminator Big Dye 75% of reads  
Chemistry: Dye-terminator Big Dye 75% of reads  
Assembly program: Phrap, version 0.990329  
Consensus quality: 131675 bases at least Q40  
Consensus quality: 135730 bases at least Q30  
Consensus quality: 138146 bases at least Q20  
Estimated insert size: 134201; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-gel estimation  
Quality coverage: 5.1x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see <http://www.hgsc.bcm.tmc.edu/docs/genbank-draft-data.html>).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 54266: contig of 54266 bp in length  
\* 54367 54366: gap of unknown length  
\* 102107 102107: contig of 47741 bp in length  
\* 102208 102207: gap of unknown length  
\* 132268 132268: contig of 30061 bp in length  
\* 132369 132368: gap of unknown length  
\* 134381 134381: contig of 2013 bp in length  
\* 134382 134381: gap of unknown length  
\* 134482 134481: gap of unknown length  
\* 136150 136149: contig of 1668 bp in length  
\* 136250 136249: gap of unknown length  
\* 137923 137922: contig of 1673 bp in length  
\* 138023 138022: gap of unknown length  
\* 139064 139064: contig of 1042 bp in length.

FEATURES  
Location/Qualifiers  
1..139064  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-100E14"

BASE COUNT 43158 a 27043 c 26771 g 41475 t 617 others  
ORIGIN

Query Match 87.6%: Score 18.4; DB 2; Length 139064;  
Best Local Similarity 95.0%: Pred. No. 6.3e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 tgaagaaggaagctaaagt 21  
||||| |||||||  
Db 4584 TGAAGAAGCAAGCTAAAGT 4565

RESULT 9  
AC011421 148142 bp DNA PRI 18-APR-2000  
LOCUS Homo sapiens chromosome 5 clone CTD-230305, complete sequence.  
AC011421  
AC011421.2 GI:6102634  
VERSION  
KEYWORDS HTG.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 148142)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 148142)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Direct Submission  
REFERENCE 3 (bases 1 to 148142)  
AUTHORS Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
JOURNAL DOE Joint Genome Center and Stanford Human Genome Center.  
TITLE Direct Submission  
REFERENCE 4 (bases 1 to 148142)  
AUTHORS Submitted (23-OCT-1999) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
REFERENCE 5 (bases 1 to 148142)  
AUTHORS Submitted (02-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
REFERENCE 6 (bases 1 to 148142)  
AUTHORS Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
JOURNAL On Oct 23, 1999 this sequence version replaced gi:6013521.  
COMMENT Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.9% of Sequence;  
Estimated Total Number of Errors is 0.1.  
STS Content:  
WT-6744 G05721  
WT-6988 G06342.

FEATURES  
Location/Qualifiers  
1..148142  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-230305"

BASE COUNT 50576 a 26865 c 24848 g 45853 t  
ORIGIN

Query Match 87.6%: Score 18.4; DB 9; Length 148142;  
Best Local Similarity 95.0%: Pred. No. 6.4e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgaagaaggaagctaaag 20  
||||| |||||||  
Db 40106 ATGAAGAAGCAAGCTAAAG 40125

RESULT 10  
AC091111 156048 bp DNA HTG 11-AUG-2001  
LOCUS Homo sapiens chromosome 18 clone RP11-837W2 map 18, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*; 1 ordered pieces.  
AC091111  
AC091111.3 GI:15028541  
VERSION HTG; HTGS\_PHASE2; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
KEYWORDS human.  
SOURCE Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 156048)  
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE Homo sapiens chromosome 18, clone RP11-837M2.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 156048)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,  
Barra, N., Bastien, V., Boguslavsky, L., Bouckhale, B., Brown, A.,  
Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M.,  
Collins, S., Collymore, A., Cooke, P., Deatellano, K., Dewar, K.,  
Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Glinde, S., Goyette, M., Graham, L.,  
Grand-pierre, N., Hagos, B., Heatford, A., Horton, L., Hulme, W.,  
Iliev, I., Johnson, R., Jones, C., Karatas, A., Labocque, K.,  
Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
Maclean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M.,  
McEwan, P., McKernan, K., McPheters, R., Meldim, J., Meus, L.,  
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,  
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S.,  
Severy, P., Sougnuez, C., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,  
Theodore, J., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zahoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (29-MAR-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Jul 29, 2001 this sequence version replaced gi:14389354.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RW/RepeatMasker.html

----- Genome Center -----  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIGSC  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information -----  
Center project name: LI2690  
Center clone name: 837\_M2

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 156048: contig of 156048 bp in length.

FEATURES  
source 1. 156048  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="18"  
/map="18"  
/clone="RP11-837M2"  
/clone\_11b="RP11-837M2"  
/clone\_11b="RP11-837M2"  
BASE COUNT 46624 a 30207 c 30590 g 48568 t 59 others  
ORIGIN

Query Match 87.6%; Score 18.4; DB 2; Length 156048;  
Best Local Similarity 95.0%; Pred. No. 6.4e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 tgaagaagaagctaaagt 21  
||||| |||||||  
Db 125436 TGAAAAGCAAGCTAAAGT 125455

RESULT 11  
AC023145  
LOCUS  
DEFINITION  
AC023145  
AC023145.4 GI:7344691  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 159012)  
AUTHORS Waterston, R.H.  
TITLE The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 159012)  
JOURNAL Waterston, R.H.  
REFERENCE Waterston, R.H.  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT Submitted (08-FEB-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Mar 30, 2000 this sequence version replaced gi:7235302.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WIGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H.NH0621321  
----- Summary Statistics -----  
Sequencing vector: M13; 1008  
Sequencing vector: plasmid; 08  
Chemistry: Dye-primer ET; 91% of reads  
Chemistry: Dye-terminator Big Dye; 9% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 149846 bases at least Q40  
Consensus quality: 15197 bases at least Q30  
Consensus quality: 153141 bases at least Q20  
Insert size: 184000; agarose-fp  
Insert size: 156512; sum-of-contigs  
Quality coverage: 3.44 in Q20 bases; agarose-fp  
Quality coverage: 4.11 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 26 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1282: contig of 1282 bp in length  
\* 1283 1382: gap of unknown length  
\* 2733 2832: contig of 1350 bp in length  
\* 2833 4155: gap of unknown length  
\* 4156 4255: contig of 1323 bp in length  
\* 4256 6155: gap of unknown length  
\* 6156 6255: contig of 1900 bp in length  
\* 6256 7748: gap of unknown length  
\* 7749 7848: contig of 1493 bp in length  
\* 7849 10056: gap of unknown length  
\* 10057 10156: contig of 2208 bp in length  
\* 10157 12429: gap of unknown length  
\* 12430 12529: contig of 2273 bp in length  
\* 12530 15456: gap of unknown length  
\* 15457 15556: contig of 2927 bp in length  
\* 15557 20085: gap of unknown length  
\* 20086 20185: contig of 4523 bp in length  
\* 20186 23226: gap of unknown length  
\* 23227 23326: contig of 3041 bp in length  
\* 23326: gap of unknown length

```
* 23327 26489: contig of 3163 bp in length
* 26489 26589: gap of unknown length
* 26589 29067: contig of 2478 bp in length
* 29068 29167: gap of unknown length
* 29168 33179: contig of 4012 bp in length
* 33180 33279: gap of unknown length
* 33280 37503: contig of 4224 bp in length
* 37504 37603: gap of unknown length
* 37604 43400: contig of 5797 bp in length
* 43401 43500: gap of unknown length
* 43501 48719: contig of 5219 bp in length
* 48720 48819: gap of unknown length
* 48820 54616: contig of 5797 bp in length
* 54617 54716: gap of unknown length
* 54717 60644: contig of 5927 bp in length
* 60644 60743: gap of unknown length
* 60744 69312: contig of 8569 bp in length
* 69313 69412: gap of unknown length
* 69413 81328: contig of 11816 bp in length
* 81329 91245: contig of 9917 bp in length
* 91246 91345: gap of unknown length
* 91346 104492: contig of 13147 bp in length
* 104493 104592: gap of unknown length
* 104593 116997: contig of 12405 bp in length
* 116998 127813: contig of 10716 bp in length
* 127814 127913: gap of unknown length
* 127914 141384: contig of 13471 bp in length
* 141385 159012: contig of 17528 bp in length.
* 141485 Location/Qualifiers
  1. 159012
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="8"
    /clone="RP11-621J21"
```

## FEATURES

source

```
misc_feature
  /note="assembly_name:Contig2"
  1383..2732
misc_feature
  /note="assembly_name:Contig6"
  2833..4155
misc_feature
  /note="assembly_name:Contig7"
  4256..6155
misc_feature
  /note="assembly_name:Contig8"
  6256..7748
misc_feature
  /note="assembly_name:Contig9"
  7849..10056
misc_feature
  /note="assembly_name:Contig10"
  10157..12429
misc_feature
  /note="assembly_name:Contig11"
  12530..15456
misc_feature
  /note="assembly_name:Contig12"
  15557..20085
misc_feature
  /note="assembly_name:Contig13"
  20186..23226
misc_feature
  /note="assembly_name:Contig14"
  clone_end:T7
  vector_side:left"
  23327..26489
misc_feature
  /note="assembly_name:Contig15"
  26590..29067
misc_feature
  /note="assembly_name:Contig16"
  29168..33179
misc_feature
  /note="assembly_name:Contig17"
  33280..37503
misc_feature
  /note="assembly_name:Contig18"
  37604..43400
misc_feature
  /note="assembly_name:Contig19"
  43501..48719
misc_feature
  /note="assembly_name:Contig20"
  48820..54616
misc_feature
  /note="assembly_name:Contig21"
```

```
misc_feature 54717..60643
  /note="assembly_name:Contig22"
  60744..69312
misc_feature /note="assembly_name:Contig23"
  69413..81228
misc_feature /note="assembly_name:Contig24"
  clone_end:SP6
  vector_side:left"
  81329..91245
misc_feature /note="assembly_name:Contig25"
  91346..104492
misc_feature /note="assembly_name:Contig26"
  104593..116997
misc_feature /note="assembly_name:Contig27"
  117098..127813
misc_feature /note="assembly_name:Contig28"
  127914..141384
  /note="assembly_name:Contig29"
  141485..159012
  /note="assembly_name:Contig30"
  159012..17528
BASE COUNT 53129 a 27457 c 27795 g 48121 t 2510 others
ORIGIN
```

```
Query Match 87.6%; Score 18.4; DB 2; Length 159012;
Best Local Similarity 95.0%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 tgaagaagaaagctaaagt 21
  |||||
Db 43587 TGAAGAAGCAATCTAAAGT 43606
```

```
RESULT 12
AP001493/c
LOCUS
DEFINITION
  Homo sapiens chromosome 18 clone RP11-779H19 map 18q21, WORKING
  DRAFT SEQUENCE, 23 unordered pieces.
ACCESSION
  AP001493.2 GI:8117351
VERSION
  HTG: HTGS_PHASE1; HTG: DRAFT.
KEYWORDS
  Homo sapiens DNA, clone:RP11-779H19.
SOURCE
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 161624)
  Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
  Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  Homo sapiens 161,624 genomic DNA of 18q21
  Published Only in Database (2000) In press
  2 (bases 1 to 161624)
  Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
  Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  Direct Submission
  Submitted (17-MAR-2000) Masahira Hattori, The Institute of Physical
  and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
  Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
  Japan (E-mail:hattori@gs.c.riken.go.jp,
  URL:http://npg.gsc.riken.go.jp/, Tel:81-42-778-9923,
  Fax:81-42-778-9924)
  On May 30, 2000 this sequence version replaced gi:7288187.
  ----- Genome Center
  Center: RIKEN Genomic Sciences Center(GSC)
  Center code: RIKEN
  Web site: http://npg.gsc.riken.go.jp/
  Contact: hattori@gs.c.riken.go.jp
  ----- Project Information
  Center project name: Humdrat18
  Center clone name: RP11-779H19
  ----- Summary Statistics
  Sequencing vector: PCR products: 100% of reads
  Chemistry: Dye-terminator ET-amersham: 100% of reads
  Assembly program: Phrap: version 0.990329
```

## COMMENT



Consensus quality: 150399 bases at least Q40  
Consensus quality: 155891 bases at least Q30  
Consensus quality: 158205 bases at least Q20  
Insert size: 159424; sum-of-ctrls  
Quality coverage: 5.50x in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 32754 contig of 32754 bp in length  
32855 48719 contig of 15865 bp in length  
48820 64088 contig of 15269 bp in length  
64189 76256 contig of 12068 bp in length  
76357 86295 contig of 9939 bp in length  
86396 94552 contig of 8157 bp in length  
94653 102595 contig of 7943 bp in length  
102596 110667 contig of 7972 bp in length  
110768 116555 contig of 5788 bp in length  
116656 122099 contig of 5444 bp in length  
122200 125461 contig of 3262 bp in length  
125562 131077 contig of 5516 bp in length  
131178 135924 contig of 4747 bp in length  
136025 138309 contig of 2285 bp in length  
138410 140832 contig of 2423 bp in length  
140933 144613 contig of 3681 bp in length  
144714 147380 contig of 2667 bp in length  
147481 150820 contig of 3340 bp in length  
150921 154114 contig of 3194 bp in length  
154215 156288 contig of 2074 bp in length  
156389 158482 contig of 2094 bp in length  
158583 160542 contig of 1860 bp in length  
160543 161624 contig of 1082 bp in length

Sequence updated (26-May-2000).

NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 32754 contig of 32754 bp in length  
32755 32854: gap of 100 bp  
32855 48719: contig of 15865 bp in length  
48720 48819: gap of 100 bp  
48820 64088: contig of 15269 bp in length  
64089 64188: gap of 100 bp  
64189 76256: contig of 12068 bp in length  
76257 76356: gap of 100 bp  
76357 86295: contig of 9939 bp in length  
86296 86395: gap of 100 bp  
86396 94552: contig of 8157 bp in length  
94553 94652: gap of 100 bp  
94653 102595: contig of 7943 bp in length  
102596 102695: gap of 100 bp  
102696 110667: contig of 7972 bp in length  
110668 110767: gap of 100 bp  
110768 116555: contig of 5788 bp in length  
116556 116655: gap of 100 bp  
116656 122099: contig of 5444 bp in length  
122100 122199: gap of 100 bp  
122200 125461: contig of 3262 bp in length  
125462 125561: gap of 100 bp  
125562 131077: contig of 5516 bp in length  
131078 131177: gap of 100 bp  
131178 135924: contig of 4747 bp in length  
135925 136024: gap of 100 bp  
136025 138309: contig of 2285 bp in length  
138310 138409: gap of 100 bp

\* 138410 140832: contig of 2423 bp in length  
\* 140833 140932: gap of 100 bp  
\* 140933 144613: contig of 3681 bp in length  
\* 144614 144713: gap of 100 bp  
\* 144714 147380: contig of 2667 bp in length  
\* 147381 147380: gap of 100 bp  
\* 147481 150820: contig of 3340 bp in length  
\* 150821 150920: gap of 100 bp  
\* 150921 154114: contig of 3194 bp in length  
\* 154115 154214: gap of 100 bp  
\* 154215 156288: contig of 2074 bp in length  
\* 156289 156388: gap of 100 bp  
\* 156389 158482: contig of 2094 bp in length  
\* 158483 158582: gap of 100 bp  
\* 158583 160442: contig of 1860 bp in length  
\* 160443 160542: gap of 100 bp  
\* 160543 161624: contig of 1082 bp in length.

#### FEATURES

source

misc\_feature  
1. 32754  
/note="assembly\_fragment"  
misc\_feature  
32855. 48719  
/note="assembly\_fragment clone\_end:77 vector\_slide:left"  
misc\_feature  
48820. 64088  
/note="assembly\_fragment"  
misc\_feature  
64189. 76256  
/note="assembly\_fragment"  
misc\_feature  
76357. 86295  
/note="assembly\_fragment"  
misc\_feature  
86396. 94552  
/note="assembly\_fragment"  
misc\_feature  
94653. 102595  
/note="assembly\_fragment"  
misc\_feature  
102696. 110667  
/note="assembly\_fragment"  
misc\_feature  
110768. 116555  
/note="assembly\_fragment"  
misc\_feature  
116556. 122099  
/note="assembly\_fragment"  
misc\_feature  
122200. 125461  
/note="assembly\_fragment"  
misc\_feature  
125562. 131077  
/note="assembly\_fragment"  
misc\_feature  
131178. 135924  
/note="assembly\_fragment"  
misc\_feature  
136025. 138309  
/note="assembly\_fragment clone\_end:SP6 vector\_slide:left"  
misc\_feature  
138410. 140832  
/note="assembly\_fragment"  
misc\_feature  
140933. 144613  
/note="assembly\_fragment"  
misc\_feature  
144714. 147380  
/note="assembly\_fragment"  
misc\_feature  
154215. 156288  
/note="assembly\_fragment"  
misc\_feature  
156389. 158482  
/note="assembly\_fragment"  
misc\_feature  
158583. 160442  
/note="assembly\_fragment"  
misc\_feature  
160543. 161624  
/note="assembly\_fragment"  
BASE COUNT 49551 a 30115 c 30590 g 49168 t 2200 others  
ORIGIN

Query Match 87.6%; Score 18.4; DB 2; Length 161624;  
Best Local Similarity 95.0%; Pred. No. 6.4e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 tgaagaagaagctaaagt 21  
||||| |||||||  
Db 38706 TCAAAAGCAAGCTAAAGT 38687

## RESULT 13

AC026923/c

LOCUS AC026923 164652 bp DNA HTG 20-APR-2000  
DEFINITION Homo sapiens clone RP11-725F23, WORKING DRAFT SEQUENCE, 13  
unordered pieces.

ACCESSION

AC026923.2 GI:7622414

KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE

human.

Homo sapiens

REFERENCE

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 164652)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

2 (bases 1 to 164652)

Unpublished

REFERENCE

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,  
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S.,  
Dodge, S., Domno, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
Levine, R., Lieu, G., Liu, G., Locke, K., Macdonald, P., Margulis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,  
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tessaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zahoun, J., Zimmer, A. and Zody, M.

Young, G., Zahoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 20, 2000 this sequence version replaced gi:7328788.

All repeats were identified using RepeatMasker:

Smt, A.F.A. &amp; Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

Project Information

Center project name: L7522

Center clone name: 725\_F\_23

Summary Statistics

Sequencing vector: M13; W77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 156378 bases at least Q40

Consensus quality: 160493 bases at least Q30

Consensus quality: 162306 bases at least Q20

Insert size: 14500; agarose-fp

Insert size: 163452; sum-of-contigs

Quality coverage: 5.0 in Q20 bases; agarose-fp

Quality coverage: 4.4 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1543: contig of 1543 bp in length  
\* 1544 1643: gap of 100 bp  
\* 1644 3512: contig of 1869 bp in length  
\* 3513 3612: gap of 100 bp  
\* 3613 6325: contig of 2713 bp in length  
\* 6326 6425: gap of 100 bp  
\* 6426 12232: contig of 5807 bp in length  
\* 12233 12332: gap of 100 bp  
\* 12333 17689: contig of 5357 bp in length  
\* 17690 17789: gap of 100 bp  
\* 17790 23842: contig of 6053 bp in length  
\* 23843 23942: gap of 100 bp  
\* 23943 30454: contig of 6512 bp in length  
\* 30455 30554: gap of 100 bp  
\* 30555 39019: contig of 8465 bp in length  
\* 39020 39119: gap of 100 bp  
\* 39120 52290: contig of 13171 bp in length  
\* 52291 52390: gap of 100 bp  
\* 52391 65949: contig of 13559 bp in length  
\* 65950 66049: gap of 100 bp  
\* 66050 84406: contig of 18357 bp in length  
\* 84407 84506: gap of 100 bp  
\* 84507 109983: contig of 25477 bp in length  
\* 109984 110083: gap of 100 bp  
\* 110084 164652: contig of 54569 bp in length.

FEATURES  
source  
1..164652  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="RP11-725F23"  
/clone\_lib="RP11-725F23"

misc\_feature  
1..1543  
/note="assembly\_fragment"  
misc\_feature  
1644..3512  
/note="assembly\_fragment"  
misc\_feature  
3613..6325  
/note="assembly\_fragment"  
misc\_feature  
6426..12232  
/note="assembly\_fragment"  
misc\_feature  
12333..17689  
/note="assembly\_fragment"  
misc\_feature  
17790..23842  
/note="assembly\_fragment"  
misc\_feature  
23943..30454  
/note="assembly\_fragment"  
misc\_feature  
30555..39019  
/note="assembly\_fragment"  
misc\_feature  
39120..52290  
/note="assembly\_fragment"  
misc\_feature  
52391..65949  
/note="assembly\_fragment"  
misc\_feature  
66050..84406  
/note="assembly\_fragment"  
misc\_feature  
84507..109983  
/note="assembly\_fragment"  
misc\_feature  
110084..164652  
/note="assembly\_fragment"  
BASE COUNT 51881 a 28406 c 28375 g 54786 t 1204 others  
ORIGIN



```

misc-feature      14304. .147902
                  /note="assembly-fragment"
misc-feature      148003. .152506
                  /note="assembly-fragment"
misc-feature      152607. .155699
                  /note="assembly-fragment"
misc-feature      155800. .158705
                  /note="assembly-fragment"
misc-feature      158806. .162039
                  /note="assembly-fragment"
misc-feature      162140. .164114
                  /note="assembly-fragment"
misc-feature      164215. .166502
                  /note="assembly-fragment"
misc-feature      166603. .169081
                  /note="assembly-fragment"
misc-feature      169182. .170270
                  /note="assembly-fragment"
BASE COUNT      51510 a 32818 c 32593 g 51749 t 1600 others
ORIGIN

```

```

Query Match      87.6%; Score 18.4; DB 2; Length 170270;
Best Local Similarity 95.0%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 tgaagaagaactaaagt 21
|||||
Db 84709 TCNAAAAGCAACCTAAAGT 84690

```

```

RESULT 15
AC090405          AC090405 170413 bp DNA HTG 27-MAY-2001
LOCUS             Homo sapiens chromosome 18 clone RP11-779H19 map 18, WORKING DRAFT
DEFINITION        SEQUENCE, 12 unordered pieces.
ACCESSION         AC090405
VERSION           AC090405.2 GI:14210571
KEYWORDS          HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTROP.
SOURCE            human.
ORGANISM          Homo sapiens
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE         Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS           Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
TITLE             1 (bases 1 to 170413)
JOURNAL           Birren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE         Homo sapiens chromosome 18, clone RP11-779H19
AUTHORS           Unpublished
                  2 (bases 1 to 170413)
                  Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
                  Barina,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
                  Canarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
                  Collamore,A., Cooke,P., Deatrellano,K., Dewar,K., Diaz,J.S.,
                  Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
                  Gardyna,S., Ginde,S., Goette,M., Graham,L., Grand-Pierre,N.,
                  Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
                  Jones,C., Karatas,A., Lacroque,K., Lamazares,R., Landers,T.,
                  Lehoczyk,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
                  McQuatters,R., Mathews,C., McCarthy,M., McEwan,P., McKernan,K.,
                  Murphy,T., Naylor,J., Nguyen,C., Nordu,C., Norman,C.H.,
                  O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
                  Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
                  Rhinback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M.,
                  Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,
                  Sougnier,C., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
                  Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
                  Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
                  Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
                  Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
JOURNAL           Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
COMMENT           Research, 320 Charles Street, Cambridge, MA 02141, USA
                  On May 27, 2001 this sequence version replaced gi:12958041.

```

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L12714

Center clone name: 779\_H.19

----- Summary Statistics

Sequencing Vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 165803 bases at least Q40

Consensus quality: 167767 bases at least Q30

Consensus quality: 168509 bases at least Q20

Insert size: 157000; agarose-fp

Insert size: 169313; sum-of-contigs

Quality coverage: 8.9 in Q20 bases; agarose-fp

Quality coverage: 8.3 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1 27912: contig of 27912 bp in length
* 27913 28012: gap of 100 bp
* 28013 28759: contig of 747 bp in length
* 28760 28859: gap of 100 bp
* 28860 29596: contig of 737 bp in length
* 29597 29696: gap of 100 bp
* 29697 30720: contig of 1024 bp in length
* 30721 30820: gap of 100 bp
* 30821 31522: contig of 702 bp in length
* 31523 31622: gap of 100 bp
* 31623 32368: contig of 746 bp in length
* 32369 32468: gap of 100 bp
* 32469 33330: contig of 862 bp in length
* 33331 33430: gap of 100 bp
* 33431 34637: contig of 1207 bp in length
* 34638 34737: gap of 100 bp
* 34738 38948: contig of 4211 bp in length
* 38949 39048: gap of 100 bp
* 39049 43570: contig of 4522 bp in length
* 43571 43670: gap of 100 bp
* 43671 162170: contig of 118500 bp in length
* 162171 162270: gap of 100 bp
* 162271 170413: contig of 8143 bp in length.
Location/Qualifiers
1. 170413
/oranism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-779H19"
/clone_lib="RPCI-11 Human Male BAC"
1. 27912
/note="assembly-fragment
clone_end:SP6
vector_side:left"
28013. 28759
/note="assembly-fragment"
28860. 29596
/note="assembly-fragment"
29697. 30720
/note="assembly-fragment"
30821. 31522
misc-feature

```

```

TITLE
JOURNAL
COMMENT

```

```

misc_feature      /note="assembly_fragment"
31623..32368
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
32469..33330
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
33431..34637
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
34738..38948
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
39049..43570
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
43671..162170
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
162271..170413
/note="assembly_fragment"
clone_end:T7
vector_side:right"

```

```

BASE COUNT      50853 a 32105 c 32644 g 53709 t 1102 others
ORIGIN

```

```

Query Match      87.6%; Score 18.4; DB 2; Length 170413;
Best Local Similarity 95.0%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      2 tgaagaaggaagcctaaagt 21
|||||
Db 164562 TGAAGAAGCAAGCTAAAAGT 164581

```

```

Search completed: March 20, 2002, 01:59:19
Job time: 8555 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2002, 23:03:34 ; Search time 583.26 Seconds  
(without alignments)  
44.535 Million cell updates/sec

Title: US-09-004-395-2

Perfect score: 1719

Sequence: 1 MKRKASILFFLLSTVLFQAQ.....FQNFVEKIESEKPEESSPKN 341

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1701	99.0	344	2	C70183 flagellar filament
2	361.5	22.0	350	2	D71348 flagellar filament
3	331.5	19.3	320	1	A43824 periplasmic flagel
4	202.5	11.8	301	2	A33814 flagellar filament
5	121.5	7.1	242	2	H71297 probable outer mem
6	112	6.5	4152	2	T31102 filamentous hemag
7	109	6.3	444	1	I40492 gntk-type transcri
8	108	6.3	580	1	F64385 hypothetical prote
9	106.5	6.2	695	2	G64327 H+-transporting AT
10	106	6.2	1181	2	C82500 icmf-related prote
11	105.5	6.1	705	2	A41322 N-acetylmuramoyl-l
12	104.5	6.1	1119	2	B70126 surface-located me
13	104.5	6.1	4919	2	T31105 hypothetical prote
14	104	6.1	499	2	D69735 xyulose kinase xy
15	103	6.0	650	2	S14949 dnak-type molecula
16	102.5	6.0	914	2	JC5574 inter-alpha-trypsi
17	102	5.9	241	2	A71298 probable flagell
18	101.5	5.9	613	2	JC6033 mosquitoicidal prot
19	101.5	5.9	622	2	A64494 hypothetical prote
20	101.5	5.9	627	2	B64710 site-specific DNA-
21	101.5	5.9	676	2	A45515 dnak-type molecula
22	101	5.9	1060	2	S06286 major merocrite su
23	101	5.9	2166	2	G70163 hypothetical prote
24	100.5	5.8	318	2	E71240 hypothetical prote
25	100	5.8	651	2	A96781 unknown protein F9
26	100	5.8	701	2	H71874 glycine-tRNA L1ga
27	99.5	5.8	401	2	A69368 hypothetical prote
28	99.5	5.8	907	2	S54353 inter-alpha-trypsi
29	99	5.8	1408	2	T45039 hypothetical prote

30	98.5	5.7	419	2	A56153 46K surface antige
31	98.5	5.7	644	2	S14950 dnak-type molecula
32	98.5	5.7	651	2	T45517 heat shock protein
33	98	5.7	635	2	S61175 probable membrane
34	98	5.7	707	2	S57157 hypothetical prote
35	98	5.7	913	2	D82885 multiple banded an
36	97.5	5.7	425	2	C64567 fucosyltransferase
37	97.5	5.7	721	2	A33319 microtubule-associ
38	97	5.6	564	2	T00135 hypothetical prote
39	97	5.6	1136	2	T40355 hypothetical prote
40	97	5.6	3848	2	T17414 TIPC protein - sll
41	96	5.6	1875	2	S38173 myosin-like protei
42	95.5	5.6	634	2	T01408 inclusion protein
43	95.5	5.6	647	2	T45522 heat shock protein
44	95.5	5.6	651	2	S46302 dnak-type molecula
45	95.5	5.6	651	2	T48271 dnak-type molecula

## ALIGNMENTS

RESULT	1	Query Match	Score	1701:	DB 2:	Length	344:
C70183	flagellar filament outer layer protein (flaA) homolog - Lyme disease spirochete	Best Local Similarity	98.8%	Pred. No. 5.8e-116;			
C:Species: Borrelia burgdorferi (Lyme disease spirochete)		Matches 337;	Conservative	2;	Mismatches	2;	Indels 0;
C:Date: 13-Feb-1998 #sequence-revision 13-Feb-1998 #ext-change 02-Sep-2000							
C:Accession: C70183							
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh							
son, D.; Peterson, J.; Kervatage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu							
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.							
Nature 390, 580-586, 1997							
A:Authors: Smith, H.O.; Venter, J.C.							
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.							
A:Reference number: A70100; M01D:98065943							
A:Accession: C70183							
A:Status: preliminary; nucleic acid sequence not shown; translation not shown							
A:Molecule type: DNA							
A:Residues: 1-344 <L>							
A:Cross-references: GB:AE001168; GB:AE000793; NID:92688598; PIDN:AA67025.1; PID:9268							
A:Experimental source: strain B31							
C:Superfamily: spirochete flagellar sheath protein							
QY	1 MKRKASILFFLLSTVLFQAQETDGLAEGSKRAEPGELVDFAEIARDPSSSTRDLTYVD 60						
DB	4 MKRKASILFFLLSTVLFQAQETDGLAEGSKRAEPGELVDFAEIARDPSSSTRDLTYVD 63						
QY	61 YVSGASGIKPEDMVVDLGINNMSVLLTPSARLQAVKNSVAVPAVAKSESKRYAGDTI 120						
DB	64 YVSGASGIKPEDMVVDLGINNMSVLLTPSARLQAVKNSVAVPAVAKSESKRYAGDTI 123						
QY	121 LGVRLVLPFSYSSAMIMPFKIPFSGESGNOFLGGLDINKTKMEIVSVSLGEYI 180						
DB	124 LGVRLVLPFSYSSAMIMPFKIPFSGESGNOFLGGLDINKTKMEIVSVSLGEYI 183						
QY	181 DLVLFEDNMGEMAYSMGTLKFGNADLIMSNNYIPNISRIKIDQVNPYPLASSKMR 240						
DB	184 DLVLFEDNMGEMAYSMGTLKFGNADLIMSNNYIPNISRIKIDQVNPYPLASSKMR 243						
QY	241 FKAFRVSKSHSKVKNFIFVYKDLRVLYDKLSVSDSDISESVFKYETSGTESLRKLK 300						
DB	244 FKAFRVSKSHSKVKNFIFVYKDLRVLYDKLSVSDSDISESVFKYETSGTESLRKLK 303						
QY	301 AHETFKRVLTKRKISIAEGSFQNFVEKIESEKPEESSPKN 341						
DB	304 AHETFKRVLTKRKISIAEGSFQNFVEKIESEKPEESSPKN 344						
RESULT	2						

071348  
flagellar filament outer layer protein flaA-1 precursor - syphilis spirochete  
N/Alternate names: 37k endoflagellar sheath protein FlaA; minor endoflagellar protein; P  
C/Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C/Date: 24-Jul-1998 #sequence accession 24-Jul-1998 #text\_change 17-Mar-2000  
C/Accession: D71348; A40160; A37187; B27559; A32351  
R/Reaser: C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McCo  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A>Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A/Reference number: A71250; MUID:98332770  
A/Accession: D71348  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-350 <COL>  
A:Cross-references: GB:AE001206; GB:AE005520; NID:g3322514; PIDN:AA65235.1; PID:g332251  
A:Experimental source: strain Nichols  
R:Isacs, R.D.; Radolf, J.D.  
Infect. Immun. 58, 2025-2034, 1990  
A>Title: Expression in Escherichia coli of the 37-kilodalton endoflagellar sheath protei  
A/Reference number: A40160; MUID:90307197  
A/Accession: A40160  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-29 <ISAL>  
A:Cross-references: GB:M63142; GB:M27915; NID:g155058  
R:Isacs, R.D.; Hanke, J.H.; Guzman-Verduzco, L.M.; Newport, G.; Agabian, N.; Morgard, M  
Infect. Immun. 57, 3403-3411, 1989  
A>Title: Molecular cloning and DNA sequence analysis of the 37-kilodalton endoflagellar  
A/Reference number: A37187; MUID:90035409  
A/Accession: A37187  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 21-350 <ISA2>  
A:Cross-references: GB:M26525; NID:g155056  
R:Blanco, D.R.; Champion, C.I.; Miller, J.N.; Lovett, M.A.  
Infect. Immun. 56, 168-175, 1988  
A>Title: Antigenic and structural characterization of Treponema pallidum (Nichols strain  
A/Reference number: A91753; MUID:88085470  
A/Accession: B27559  
A:Molecule type: protein  
A:Residues: 21-28, 'X', 30-36, 'G', 38-40 <BLA>  
R:Norris, S.J.; Chaston, N.W.; Cook, R.G.; Fuentes, M.D.; Lamberger, R.J.  
J. Bacteriol. 170, 4072-4082, 1988  
A>Title: Antigenic relatedness and N-terminal sequence homology define two classes of p  
A/Reference number: A91882; MUID:88314903  
A/Accession: A32351  
A:Molecule type: protein  
A:Residues: 21-48, 'R', 50-51 <NOR>  
C/Genetics:  
C:Gene: TP0249  
C:Superfamily: spirochete flagellar sheath protein  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-350/Product: flagellar filament outer layer protein flaA-1 #status experimental <MA

```

Oy      163  IKTKEIVSVYSGEYELDELVEFEDNGMGMYAAVSMGLTKFKGNADILMSPNYIPNLS 222
      178  VGVLSIAVNTYGMNYPHGLTYVMRDQGEVHRIFMGITLFDMSKELVWNNPSTISDVS 237
Oy      223  RIHKDVENYPLASSKMFKAFRVSKSSHSSVKYKNEFIYVKDLRVLYDKLSVDSIDSDSE 282
      238  REVR-LYEVYPASTPHVVEGEMVTRDAANHAAGDVGVEFKDKVITIKYDKAVLSTVRFADE 296
Oy      283  SVFKVYFTSGTESLRKLKAHETFKRVLKLREKISIAESFQNFPEKITESEK 333
      297  DLW-----GIQARREAE-----RRKVEYARPGQOOVLRYTIEQEK 330
RESULT      3
A:3824
periplasmic flagellar sheath protein precursor - Treponema hyodysenteriae
C:Species: Treponema hyodysenteriae
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jul-1999
C:Accession: A43824
R.Koopman, M.B.; de Laeuw, O.S.; van der Zelfst, B.M.; Kusters, J.G.
Infect. Immun. 60, 2920-2925, 1992
A:Title: Cloning and DNA sequence analysis of a Serpulina (Treponema) hyodysenteriae
A:Reference number: A43824; MUID:92307926
A:Accession: A43824
A:Molecule type: DNA
A:Residues: 1*320 <KOO>
A:Cross-references: EMBL:X63006; NID:9433523; PIDN:CAA44735.1; PID:9433524
C:Comment: The flagella of this spirochete are reported to be enclosed completely with
C:Genetics:
A:Gene: flaA
C:Superfamily: spirochete flagellar sheath protein
C:Keywords: flagellum; periplasmic space
E:1-19/Domain: signal sequence #status predicted <SIG>
F:20-320/Product: periplasmic flagellar sheath protein #status experimental <MAT>

```

```

Query March 19 13: Score 331.5: DB 1: Length 320;
Best Local Similarity 28.88; Pred. No. 9,4e-17;
Matches 90; Conservative 63; Mismatches 129; Indels 31; Gaps

QY 9 LEFLSTVPLFQNEFDGLAESGSKRAPEGLVLDPAELARDPSSRTLDLTNNYDYSYGASG 68
      ||:::| | | | |::||| | | | | |
DB 4 LEVYITSTIFIAASAYGLTNSI-----LIDPA-----LTGMDNLDAGSGD 43

QY 69 IYKPEDWYVDIGINNWSVLLTPSARLQAYKNSVAPAVYKSESKRYAGDTILGVRVLP 128
      : : | | | | | | | | | | | | : : : | | | | |
DB 44 TNEVVPVAVENLYNNWYVWLNESARLTENRNRSYVTVNDVSKNNGNAGMEAGKVLGYRVHP 103

QY 129 SYSSQS-AMINPPKIPFIYSGESNOPL-CKGLIDNLIKMKKEIKYSVSLGEIDLEVL 186
      : : : : : : : : : : : : : : : : | | | | |
DB 104 LAANMYSYLVLPVYELEMYGGAODTKYTECKGVIHVWGEIKTSSWVYGRNLLISYFNL 163

QY 187 EDMNGMEVAYSMGTLEKFKGNADLTWSNPNIIPISRIIKDDVPVYPLASSMKRKARV 246
      : : | | | | | | | | | | | | : : : : :
DB 164 QNEBELKSYMGVYNGNRQVMEKRELPLVNRORVLYRE-PLYPRNIPYKLDISLGF 222

QY 247 SKSHSSKYVKNFIYVKDLRVLYDKLSYIDSIDISESVKRYEYTSGETSLRKLKAHETFK 306
      : : | | | | | | | | | | | | : : : : :
DB 222 YRTKDTKGGDFYVVKDVTLEYDVVVDFEEDIDDEATWQLK---TENDRK-QAIES-- 276

QY 307 RVLKLRKISIAE 319
      : : | | : :
DB 277 --ARIREQAEIRD 287

RESULT 4
A32814
flagellar filament surface antigen - Spirochaeta aurantia
C:Species: Spirochaeta aurantia
C>Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 07-Feb-1997
C:Accession: A32814
J:Brighamsha, B.: Greenberg, E.P.
J: Bacteriol. 171, 1692-1697, 1989

```



A:Title: Cloning and sequence analysis of flaA, a gene encoding a Spirochaeta aurantia  
A:Reference number: A32814; MUID:89155480  
A:Accession: A32814  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-301 <BRA>  
C:Keywords: surface antigen

Query Match 11.8%; Score 202.5; DB 2; Length 301;  
Best Local Similarity 31.0%; Pred. No. 1.9e-07;  
Matches 57; Conservative 32; Mismatches 68; Indels 27; Gaps 7;

Db 1 MKRKASILEFLFPAQETDGLAEGSKRAEPGLVDFEALARPSSRLDITNYVD 60  
1 MKR-----FFAIGALAEFGNSGAFPEQA-----TLIDFSKIVGE-GNTGLAPPTTD 47  
Qy 61 YVYSGASGIKPED---MVVDLGINMNSVLLTPSARLQAVKNSVPAVAVKSEKRYAG 117  
48 YSRQAGSA-YSAEDKAMKISLAIPSEIELASSQTEVENQTLSTLYAAVYKODAAHYGG 106  
Db 118 DTLGVRVLFPSY-SSAMIMPPKIPY-----SGESGNOFLGKGLDNIKTMK 167  
107 ETVMGVRHPPSPGINSFAVIKPPETIPAYATLGDATQANAVAGGQDFGEGVLKNVGVIK 166  
Qy 168 EIKV 171  
167 SIQI 170

RESULT 5  
H71297  
probable outer membrane protein - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C>Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999  
C:Accession: H71297  
R:Praser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
rson, J.; Khakh, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A:Reference number: A71250; MUID:98332770  
A:Accession: H71297  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-242 <COL>  
A:Cross-references: GB:AE001240; GB:AE000520; NID:g3322952; PIDN:AAC65636.1; PID:g332296  
A:Experimental source: strain Nichols  
C:Genetics:  
A:Gene: TP0663

Query Match 7.1%; Score 121.5; DB 2; Length 242;  
Best Local Similarity 22.0%; Pred. No. 0.1;  
Matches 54; Conservative 31; Mismatches 105; Indels 55; Gaps 8;

Qy 48 PSSTRDLITNYVDY---YSGAGIYKPEDMVVDLGINM---SVLLTPSARLQAVKN 100  
21 PLAORSKVYQAVYFIDDFDGS-----EDQGL-AMRAAGSKRTFGFPILKYEG 70  
Qy 101 SVAPAVAVKSEKRYAGDTLIGVRVLFPSYSSSAMIMP-----PKIIPYSGSGNOFL 155  
71 MPQAVRAGSWGCKDKKARITGECFKNRGNMMLDILPKGSDYEIPLRGVSG---- 126  
Qy 156 GKGLDNIKTMKEIKVSVSLGYEIDLEVLFEPMNGMEYAVSMGTLEFKGNADLWISNP 215  
127 -----FDVWVWAGVQYSLVALVROCTGVRHLLIGNLDFQGNKKNLSVGP 173  
Qy 216 YIPNISRIIKDVPNPLASSKMRFAFVSKSHSKVKNFTFYVKDLRYLKLVS 275  
174 HIQTS-----RYLGSQHLSEFVGFRIRTXPSERVDVY-----VFEQFALA 217  
Qy 276 DSDID 280

Db 218 NMHID 222

RESULT 6  
T31102  
filamentous hemagglutinin 1 - Haemophilus ducreyi

C:Species: Haemophilus ducreyi  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T31102  
R:Ward, C.K.; Lumley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.  
J. Bacteriol. 180, 6013-6022, 1998  
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.  
A:Reference number: 220984; MUID:99030326  
A:Accession: T31102  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4152 <MAR>  
A:Cross-references: EMBL:AF057695; NID:g32929017; PID:g32929018; PIDN:AAC79757.1  
C:Genetics:  
A:Gene: hspA1

Query Match 6.5%; Score 112; DB 2; Length 4152;  
Best Local Similarity 22.8%; Pred. No. 28;  
Matches 84; Conservative 44; Mismatches 133; Indels 108; Gaps 15;

Qy 4 KAKSILEFLFPAQETDGLAEGSKRAEPGLVDFEALARPSSRLDITNYVDY 63  
12 KVNCLVPVAVENIKSASGSSGSSNSKIAEDQ-----EEDPSLACSL----- 55  
Db 64 SGAGIYKPEDMVVDLGINMNSVLLTPSARLQAVKNSVPAVAVKSEKRYAG 97  
56 -----PLSSSTHLGHNHSPKVEKGSLSVLLSLMPTPLAQONTAEALNGKY 107  
Qy 98 VKNVPAVAVKSEKRYAGDTLIGVRVLFPSYSSSAMIMPPIKIP-----FYSG-ESG 151  
108 VDSHSSSTRIEQKTDNSKGIYVELANEDVGSDNRKKEFNIPSAFNNSTEST 167  
Db 152 NOFLGKGLDNIKTMKEIKVSVSLGYEIDLEVLFEPMNGMEYAVSMGTLEFKG- WADLI 210  
168 SOLVGK-LHANIQLQKVK-----LILNYGTGHNESNIQALFVAGKADLI 213  
Qy 211 WSNPIYIPNISRIIKDVPNPLASSKMRKARVKS-----HSSKYNFTFYVKDLRY 266  
214 IVNNGTTLNGVKRTIND-----RRVSTSDIIPR--ENGLLSVRNCKV 256  
Qy 267 LYDLKLSV-----IDSDISESVFYETSGTESLRKLAHEFP-----KRVTLKR 312  
Db 257 TIDKGVATNGLSHFEVAVARNIDQKRTTAKTENQKSVN--PANITFAAGSLATNKLTR 314  
Qy 313 EKISIAEGS 321  
Db 315 EATPISSGT 323

RESULT 7  
I40492  
gntR-type transcription regulator homolog yxd - Bacillus subtilis  
N:Alternate names: hypotheical protein 8 (arfa operon)  
C:Species: Bacillus subtilis  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: I40492; F69766; S34990  
R:Cosma, P.; Rodrigue, F.; de Ferra, F.; Grandi, G.; Perego, M.; Venema, G.; van S  
Mol. Microbiol. 8, 821-831, 1993  
A:Title: Sequence and analysis of the genetic locus responsible for surfactin synthe  
A:Reference number: I40485; MUID:93360813  
A:Accession: I40492  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-444 <RES>  
A:Cross-references: EMBL:X70356; NID:g396480; PIDN:CAA49821.1; PID:g396486  
A:Experimental source: ATCC 21332

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capano, V.; Carter, N.M.; Cho A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Hentut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F. Koster, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois, A.; Lauber, J.; Lazarovic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel Y. M.; Ogawa, K.; Ogikawa, A.; Oudaga, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, A.; Schleich, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Serot Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipet, A.; Yamamoto, K.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K. A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. A:Reference number: A6580; MUID:98044033

A:Accession: F69766

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-444 <KUN>

A:Cross-references: GB:299105; GB:299106; GB:AL009126; NID:g2632653; PID:g2632657; NID:g

A:Experimental source: strain 168

A:Genetics:

A:Gene: yocD

C:Superfamily: hypothetical protein b1439

Query Match 6.3%; Score 109; DB 1; Length 444;

Best Local Similarity 21.6%; Pred. No. 2;

Matches 75; Conservative 42; Mismatches 126; Indels 104; Gaps 15;

29 SKRPPGELVLDFAELARDPSSRLDTNYVD-----YVSGASGI--- 69

DB 73 SMSGQGP--IDFATSRPP-----DVFYLDPOHCINKAIDPYKNDLFTGPKGLPL 125

70 -----VKPEDMVLDGINNSVLL-----TPSARLQAYKN 100

126 IRLKRLATQOVFADEHIFITSGVOALSLCAMPFNGEKIAIEQPGYHLMWEQLE 185

101 SVVAVAVYKESKRYAGDTLLGVRLFPYSOSSAMIMPFKIPFSGSGNOFLKGLI 160

186 TLGTRPAIGVKRTTE--GLDIAVERLFQESTIKFPTMRFNHP-----LGSLS 233

161 DNIMTKELKSVSYSGEIDLEVLFDNMGMAYVSMGTLKFKGWADLIWSPNYIPNI 220

234 EKDQ-----ELVRLAEAYDYVLVEDQ-----LGDLENKKAD-----PLAYDL 274

221 SRIITKDVDPNPLASSKMRKFAFVSKSHSKVKNFIYVKDLRYLKLVSIDSID 280

275 SSHVI-----YLKSFSSKMPGLRGAAVLPALDTPFYA-----YKLN-DIDCSMI 321

281 SESVFKVYETSGTESLRKLAHETFK--RYLKLREKI-----STAESF 322

DB 322 SQAALEIYLKSGMYGRHKRIKIDSTKRESLRLHQAIRHRLGSGRF 368

hypothetical protein M0686 - *Methanococcus jannaschii*

C:Species: *Methanococcus jannaschii*

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 22-Oct-1999

C:Accession: F64385

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,

son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,

A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*

A:Reference number: A64300; MUID:96337999

A:Accession: F64385

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-580 <BUJ>

A:Cross-references: GB:067515; GB:L77117; NID:g1591390; PIDN:AAB9681.1; PID:g1499499

C:Genetics:

A:Map position: REV615409-613667

Query Match 6.3%; Score 108; DB 2; Length 580;

Best Local Similarity 21.7%; Pred. No. 3.4;

Matches 56; Conservative 38; Mismatches 66; Indels 98; Gaps 11;

112 SKRYAGDTLLGVRLFPYSOSSAMIMPFKIPFSGSGNOFLKGLINDIKTKMEIKV 171

DB 376 SRRYSGST-----ESKMKDKFEVSQWIEENNNI--PEVENLKI--EIQY 417

172 SVYSLEYEIDLEVLFDNMGMAYVSMGTLKFGWADLIWSPN--YIPNISRIIKD 228

DB 418 GAYSL-----KKVSSSGSKYKGVNLIFFKKPKMDFYKP-----DN 453

229 VNNYPLASSKMRKFAF-----RVSKSHSKVKNFIYVK 262

DB 454 IAYVKLEDDHIFPKGFLRNKGISNEYIDSVLKNKPTILDETNNKISKSPS-----YVK 507

263 DLRYLYDKLSVSDSID-----SESVFKVYETSGTESLRKLAHETFKRVLKL 311

DB 508 EMIEIQKNKGLSEDEAVNKKVEILGHLNEMFELIRNF--DQSLSKDEIEENFNFIEL 566

312 REKISIAEGSFQNFVEKI 329

DB 567 REKL-----ILEKI 575

RESULT 9

64327 H-transporting ATP synthase (EC 3.6.1.34) chain I - *Methanococcus jannaschii*

C:Species: *Methanococcus jannaschii*

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000

C:Accession: G64327

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak

son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannasc*

A:Reference number: A64300; MUID:96337999

A:Accession: G64327

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-695 <BUJ>

A:Cross-references: GB:067478; GB:L77117; NID:g1590958; PIDN:AAB96208.1; PID:g1590960

A:Genetics:

A:Map position: REV214151-212064

A:Start codon: GTG

C:Keywords: hydrolase

Query Match 6.2%; Score 106.5; DB 2; Length 695;

Best Local Similarity 23.5%; Pred. No. 5.7;

Matches 84; Conservative 41; Mismatches 108; Indels 125; Gaps 19;

68 GYKPPEDMVVDGINNWSVLLPSSARLQAYVNSVAVPAVYVSESKRYAGDTLLGVRLV- 126

DB 32 GIVELCDDLESEKLEDEMKTLSPSSAD--YVRN--VTSIMK-----AG-----RIID 76

127 -PPYSOSSAMIMPFKIPFSGSGNOFLKGL-----ID 161

DB 77 MESSVQKETSISKDILNPKPVEKKKVSFOEVIDYAEKVLNLSKEVDGAERISELD 136

162 NIKT--MKEIKVSV--YSLGEIDLEVLFDNMGMAYV-----SMGTLKFGWADLI 210

DB 137 NKKSKTLQKEDISYKGLGEPDLKYL-----GSGEYVFIAGSVPEKELGELK-----AELD 188

211 WSNPNYIPNIS--SRIITKDVDPNPLASSKMRKFAFVSKSHSKVKNFIYVKDLRYLVD 269

DB 189 KVADGYTIGFSGSEFEKDKIRVPIVFTYKLELVN-----LSEIRKFEFEERYDI----- 239

```

Oy      270 KLSVSDSDS-----ESVKEVETSGTSELRLK--AHHTFRVLKLRKLSI- 317
          ||::|||::|||::|||::|||::|||::|||::|||::|||
Db      240 -----SDVECTPSALSKIETSELKAIESERNSLIEKTLALQAQWEKLLAYVELLSIE 292
          ||::|||::|||::|||::|||::|||::|||::|||::|||
Oy      318 -----AEGSFQN-----FVEKIESKEPSESSP 339
          ||::|||::|||::|||::|||::|||::|||::|||::|||
Db      293 KARGAGVSGFGKTRKTYIEAMVPARDAEKAKSLIENSADGFAFVEITRPDEPEKIP 350
          ||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 10
ICMF-related protein VCA0120 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C82500
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: C82500
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
  charlson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.
  l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: C82500
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1181 <HEI>
A:Cross-references: GB:AE004353; GB:AE003853; NID:9657488; PIDN:AAF96034.1; GSPDB:GM001
C:Genetics:
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
A:Gene: VCA0120
A:Map position: 2

```

```

Query Match 5.2%; Score 106; DB 2; Length 1181;
Best Local Similarity 23.3%; Pred. No. 13;
Matches 77; Conservative 54; Mismatches 118; Indels 82; Gaps 17;

      28 GSKRAPEGEVLVDFAEL-----ARDPSSTRDLDTNYVDYVS---GASG 68
      |||  ||:  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 866 GSKPAPMTEVLVASDELKSLKSIDADPVGMAALDATKARKVLVN-ADPYTLTKRISGG 924
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 69 IYKPED-VWVDLGINMSVLLTPSARLQAYKNSVY--APAVKSESKRYAG----- 117
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 925 LPKPLDSMMAKLADSESWYV-----KQEAIKHLEVRTEDEYKTFQSKLAGRYFPNDAS 978
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 118 -DTLGVRLVLF-PS-----YSQSSAMIMPEKPIPIPSGESGNQGLGKLDINIKTMK 167
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 979 NKDVALAEFAEFPAENGTLDNFYNQQLKMFIDE-NISVASDDSAQSIIRKEVLQIKQAG 1037
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 168 EIKSVYSLGVEIDLEVEF-----DMNGEYIAYSKGTLEFKGADLIGSNP 214
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
Db 1038 KIREAFERRKGIIDVFSFEVPLSLSNKRRSVLNVGQFLAVSHCP---RENVELIWPNT 1094
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
Qy 215 -----NYP-----NISRIKIDVPENYPPLASSKMRKAFKRYKSSHKVKNPIFYK 262
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
Db 1095 LRDSAVSKAVTILPIPTQTNNSPSLQ---IQGPAFFRLLDGGVDVASQTS--VDEKFIIVD 1149
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
Qy 263 DLRLVLDKLSVSDSIDSESVFKVYETSGT 293
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
Db 1150 GSEMIY-RINMAEDANPETERLRFKSFKLST 1179

RESULT 11
A41322
N-acetyl-muramoyl-L-alanine amidase modifier precursor lytB - Bacillus subtilis
N:Alternate names: major autolysin lytC modifier lytB; probable cell wall-binding protein
C:Species: Bacillus subtilis
C:Date: 05-Jun-1992 #sequence-revision 10-Feb-1995 #text-change 15-Oct-1999
C:Accession: A41322; C47679; D69654
R:Kuroda, A.; Sekiguchi, J.
J. Bacteriol. 173, 7304-7312, 1991
A:Title: Molecular cloning and sequencing of a major Bacillus subtilis autolysin gene.
A:Reference number: A41322; MUID:92041629

```

A:Accession: A41322  
A:Molecule type: DNA  
A:Residues: 473-705 <KUN>  
A:Cross-references: GB:M01747  
R:Lazarevic, V.; Margot, P.; Soldo, B.; Karamata, D.  
J. Gen. Microbiol. 138, 1949-1951, 1992  
A:Title: Sequencing and analysis of the *Bacillus subtilis* lytRABC divergon: a regulatory  
A:Reference number: A47679; MUID:93018998  
A:Accession: C47679  
A:Molecule type: DNA  
A:Residues: 1-705 <LAZ>  
A:Cross-references: GB:M07645; NID:g143155; PIDN:AAA22580.1; PID:g143158  
A:Experimental source: strain 168  
A:Note: sequence extracted from NCBI backbone (NCBI:N:116645, NCBI:P:116648)  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabely, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, S.; Fuma, S.; Gallizi, A.; Gal  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.;  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portele  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scall  
A:Authors: Schlecht, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
akunchi, M.; Tamakoshi, A.; Tanaka, T.; Terpieta, P.; Tognoni, A.; Tosato, V.; Uchiya  
T.; Winters, P.; Wipit, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zimmstein, E.; Yoshikawa, H.; Dancho, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
A:Reference number: A69580; MUID:96044033  
A:Accession: D69654  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-705 <KUN>  
A:Cross-references: GB:291122; GB:AL009126; NID:g2636029; PIDN:CAB15580.1; PID:el1844  
A:Experimental source: strain 168  
C:Comment: This protein enhances the activity of the N-acetyluramoyl-L-alanine amidat  
C:Genetics:  
A:Gene: lytB  
A:start codon: TTG  
C:Keywords: cell wall  
F:1-75/Domian: signal sequence #status predicted <SIG>  
F:26-705/Product: N-acetyluramoyl-L-alanine amidaase modifier #status predicted <MAT>

[illegible]

RESULT 12  
B70126  
surface-located membrane protein 1 (Imp1) homolog - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C>Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 11-Jan-2000  
C:Accession: B70126  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943  
A:Accession: B70126  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1119 <LE>  
A:Cross-references: GB:AE001131; GB:AE00783; NID:92688098; PID:AA66595.1; PID:9268810  
A:Experimental source: strain B31  
C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat hc  
F:742-774/Domain: tetratricopeptide repeat homology #status atypical <TT1>  
F:775-808/Domain: tetratricopeptide repeat homology <TT2>  
F:809-842/Domain: tetratricopeptide repeat homology <TT3>  
F:843-876/Domain: tetratricopeptide repeat homology <TT4>  
F:911-943/Domain: tetratricopeptide repeat homology <TT5>  
F:944-977/Domain: tetratricopeptide repeat homology <TT6>  
F:979-1010/Domain: tetratricopeptide repeat homology #status atypical <TT7>  
F:1011-1044/Domain: tetratricopeptide repeat homology <TT8>

Query Match 6.1%; Score 104.5; DB 2; Length 1119;  
Best Local Similarity 21.7%; Pred. No. 16;  
Matches 68; Conservative 44; Mismatches 110; Indels 91; Gaps 12;

QY 98 VKNSVAAVAVKSEKRRACDITLCVRLFPSSYSSSAMIM-----PPKIRIFY 146  
DB 46 VKNSYID-----RNYKKAIS-----VAKLLDQKYPONEDIAMLTNTLAETIANSSPESKDL 96  
QY 147 SGEQNGPLG--KGLIDNKT-----MKKIKSVSYSLG-ETIDLEVLFEDMGM 192  
DB 97 QRDASNOQLDKIKGO-DTKTNVNNFNPIAFNNRIKSTITENDSDNDVDGIDEDIS 155  
QY 193 EYAVSMGLTKRKGWADLIWSPNYPINISSRIKQDVPNYPVPLASSKMPKAFRVSKSHS 252  
DB 156 EFKSKIKPEIK-----PNTMPKREDDQIIQSPNPKLSVNDQKNFLENLEKTKNSG 206  
QY 253 KVKNFIFVKKDLRVLYOKLSYISDSDISESVFKYERS-----GTESLRL----- 299  
DB 207 KNSSENILDSQKLENDQNTNLSKKNSENLKTPDNSKYSNNNNNTSLKKSINSQKE 266  
QY 300 -----KAHEFRKRVLK-----LREKISIA-EGSFQNF 325  
DB 267 SELSPQSTIIKTIKRYPSYLIKELLEYLIDDINTGRVTLGKNRLKELIKGSLSKFKQV 326  
QY 326 VEKIESKEPESS 338  
DB 327 NELIENSKNEAS 339

RESULT 13  
T31105  
hypothetical protein 2 - Haemophilus ducreyi  
C:Species: Haemophilus ducreyi  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T31105  
R:Ward, C.K.; Lumley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.  
J. Bacteriol. 180, 6013-6022, 1998  
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.  
A:Reference number: Z20984; MUID:99030326  
A:Accession: T31105  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA

A:Residues: 1-4919 <NA>  
A:Cross-references: EMBL:AF057696; NID:93929021; PID:93929023; PID:AA67961.1  
C:Genetics:  
A:Gene: lspa2

Query Match 6.1%; Score 104.5; DB 2; Length 4919;  
Best Local Similarity 19.1%; Pred. No. 1.3e+02;  
Matches 82; Conservative 58; Mismatches 119; Indels 171; Gaps 18;

QY 54 DLTVNDVYVSGASIVPEDMVVDLGT-----NNMSVLTTSARLQAV-VKNSVAAVAV 108  
DB 1001 NISNKIAVLSLGNISLNSKDDQVYNLGETIYAGNNISV-----KAHQKNDVAKMGDI 1052  
QY 109 KSESK-----RYAGDTILGVRLFPSSYSSSAMIMPP----- 141  
DB 1053 TTKTKEGQASYKLYOASNGCHGNDGSSG-----YEGDLNITKGRADDNKLTQRI 1105  
QY 142 -----KIPYSGEG--NOFLGKGLDNK-----TKKIKSVSYSLGYE 179  
DB 1106 KTIYAGRLTFKNSNAGKSEIINRGTI-NVKNKLSYSDVSFFENMOSOKVDLYTKIFE 1164  
QY 180 I--DEVLFE-----DMNGMEYAVSMGLTKRKGWADLI----- 210  
DB 1165 AKSDIELFETNGTHPYVLNFKSNNEKKRNSENTKMKFSIGDILNEALSDAPEAIEA 1224  
QY 211 -----WSNPNYIPNLSRII-----KDDVPNYP-LASSKMR 241  
DB 1225 YSSGSSSVINPVSYLALGANNNSNPHYLNTALKHLGNQWQDLDKQENIKVLRQKM 1284  
QY 242 KAFRVSKSHSKVKNFITYVDLRYLYOKLSYST-----SDSDISESVFKY-- 288  
DB 1285 EDFKDKGASKKL-----DLYPNTDEKAKIPAGIIRNGNDTISDESEDFKFKYS 1336  
QY 289 -----ETGSESLRKLAHEFRKRVLKLEKISIAEG-----SEQNFV 326  
DB 1337 FONGEWAANDGTGSDYDSTKASEKTKKVENVDHAKENIDEHKLNTIGKHEITVPGVSFENLN 1396  
QY 327 EKIESEKPEE 336  
DB 1397 NKNMDHDPDK 1406

RESULT 14  
D69735  
xylose kinase xylB - Bacillus subtilis  
C:Species: Bacillus subtilis  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: D69735  
R:Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Acevedo, V.; Ber  
C.; Bron, S.; Brouillett, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
lechi, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M  
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma  
Y., M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowa, A.; Se  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchida  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; MUID:98044033  
A:Accession: D69735  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-499 <KUN>  
A:Cross-references: GB:Z99113; GB:AL009126; NID:92634090; PID:CA613645.1; PID:926341  
A:Experimental source: strain 168  
A:Genetics:  
A:Gene: xylB

C:Superfamily: xylulokinase

Query Match 6.1%; Score 104; DB 2; Length 499;  
Best Local Similarity 20.9%; Pred. No. 5.4;  
Matches 68; Conservative 57; Mismatches 135; Indels 66; Gaps 14;

QY 30 KRAEGELVDAEELARDPSSTRLDLTNYVDVYGASG-----YVKPE---DMVVDLGI 81  
DB 141 KHEP-ELFKRAVFLPRDYRFRRTGVYHREYSDAAITLHLITRKEMSDICQIGI 199  
QY 82 NMSVLLPPSARLQAVYKNSVAVPAVYKSESKRYAGDTLLGVAVLPSPYSQSSAMIMPF 141  
DB 200 -----SADICPLVESH-----CVGSLLPHVAAKTGLE--- 229  
QY 142 KIPYSGESGNG--FLGKLLINIKTKIKSVISLGEIDLE-----VLFEDNGME 193  
DB 230 KIRVYAGGADNACGAGIGAILSSGKTLCSIGTVLLSYEEKERDPKGVHFFNGKRD 289  
QY 194 VAYSGTLEFKGMADLIMSNPNVYIPNISRIIKDVPNPYPLASSKMKFAFRYSK--SHS 251  
DB 290 SYTMGVTLAAGYS--LDWFKRTFAPNESFOLLQGVFAIPGANGLLTYPIYGERTPHA 348  
QY 252 -SKVKNFIIFYKD-----LRVLYDKLSVSDSIDID-----SESVFKYVETSG---TES 295  
DB 349 DSIKSGSLIGMDGCAHNRKHFLRAIMEGITFSLHESIELFREAGKSVHTVVSIGGAKNDT 408  
QY 296 LKRLKAHETFRVYLKREKISIAEGS 321  
DB 409 WLQMDIFNTVRVIRLENGPAMGA 434

RESULT 15

S14949  
drak-type molecular chaperone hsc-1 - tomato  
N:Alternate names: heat shock cognate protein 70  
C:Species: Lycopersicon esculentum (tomato)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Aug-1999  
C:Accession: S14949  
R:Lin. T.Y.: Duck, N.B.; Winter, J.; Folk, W.R.  
Plant Mol. Biol. 16, 475-478, 1991  
A:Title: Sequences of two hsc 70 cDNAs from Lycopersicon esculentum.  
A:Reference number: S14949; MUID: 91370894  
A:Accession: S14949  
A:Molecule type: mRNA  
A:Residues: 1-650 <LIN>  
A:Cross-references: EMBL:X54029; NID:g19255; PIDN:CAA37970.1; PID:g19256  
C:Gene: hsc-1  
C:Genetics:  
A:Function:  
A:Description: Involved in protein folding and assembling/disassembling of protein comp  
C:Superfamily: heat shock protein 70  
C:Keywords: ATP, molecular chaperone

Query Match 6.0%; Score 103; DB 2; Length 650;  
Best Local Similarity 18.8%; Pred. No. 9.2;  
Matches 68; Conservative 57; Mismatches 124; Indels 112; Gaps 13;

QY 43 ELARDPSSTRLDLTNYVDVYGASGIVKPEDMVVDLGINNSVLLTPSARLQAVYKNSV 102  
DB 61 QVALNPINTVFDAKRLIGRFSAS---VQEDMKL-----WPFKVIQPGDK----- 104  
QY 103 VAPAVY---KSESKRYAGDTLLGVAVLPSPYSQSSAMIMPF----- 142  
DB 105 --PMIVVYTKGEEKFAAEI-----SSMVLTKMKEIAEALGSLVKNVAVT 149  
QY 143 IPFYSGESGNGFL-GKGLIDNITKMEI-KSVYSLSGEYIDLE-----VLFEDNGM 192  
DB 150 VPAYFNDSQGRQATKDAAGVSGLVNVRKINEPTAAALAYGLDKKATSAKEKNVLIIFDLGG 209  
QY 193 EYAYSNGTLK-----FKGMADLIMSNPNYIPNISRIIKDVPNPYPLASSKM 239

DB 210 TPDVSLTTEEGIFEVKATAGDTHLGGEDFDNBMVNHFEFRKRKKDITGNPRALRL 269  
QY 240 RFKAFRVSKSHSSKVKNF-----FYVKDLRYLXDKLSVSI-----D 276  
DB 270 RTACERAKRTLSSTAOFTTIEIDSLYGVDPYSTITRAREELNMDLFRKCMFEVEKCLRD 329  
QY 277 SDIDSESVFKVYETSGTESLRLKAHETFRVYLKREKISIAEGSFQNFVEKIESEKPEE 336  
DB 330 AKMDKSTVHDVVLVGSTRIPKYQ-----QVAMTINFGKELCKSIMPDE 374  
QY 337 S 337  
DB 375 A 375

Search completed: March 19, 2002, 23:42:09  
Job time: 2315 sec

000000

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2002, 22:13:23 ; Search time 632 Seconds  
(without alignments)  
39.967 Million cell updates/sec

Title: US-09-004-395-2

Perfect score: 1719  
Sequence: 1 MKRRAKSLFLFLSTVLFNQ.....FQNFVEKIESEKPESSPKN 341

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_1101:\*

- 1: /SIDS2/gcgdata/geneseq/AA1980.DAT:\*
- 2: /SIDS2/gcgdata/geneseq/AA1981.DAT:\*
- 3: /SIDS2/gcgdata/geneseq/AA1982.DAT:\*
- 4: /SIDS2/gcgdata/geneseq/AA1983.DAT:\*
- 5: /SIDS2/gcgdata/geneseq/AA1984.DAT:\*
- 6: /SIDS2/gcgdata/geneseq/AA1985.DAT:\*
- 7: /SIDS2/gcgdata/geneseq/AA1986.DAT:\*
- 8: /SIDS2/gcgdata/geneseq/AA1987.DAT:\*
- 9: /SIDS2/gcgdata/geneseq/AA1988.DAT:\*
- 10: /SIDS2/gcgdata/geneseq/AA1989.DAT:\*
- 11: /SIDS2/gcgdata/geneseq/AA1990.DAT:\*
- 12: /SIDS2/gcgdata/geneseq/AA1991.DAT:\*
- 13: /SIDS2/gcgdata/geneseq/AA1992.DAT:\*
- 14: /SIDS2/gcgdata/geneseq/AA1993.DAT:\*
- 15: /SIDS2/gcgdata/geneseq/AA1994.DAT:\*
- 16: /SIDS2/gcgdata/geneseq/AA1995.DAT:\*
- 17: /SIDS2/gcgdata/geneseq/AA1996.DAT:\*
- 18: /SIDS2/gcgdata/geneseq/AA1997.DAT:\*
- 19: /SIDS2/gcgdata/geneseq/AA1998.DAT:\*
- 20: /SIDS2/gcgdata/geneseq/AA1999.DAT:\*
- 21: /SIDS2/gcgdata/geneseq/AA2000.DAT:\*
- 22: /SIDS2/gcgdata/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1719	100.0	341	20	AAV28460
2	1645	95.7	335	20	AAV20146
3	1551	90.2	315	20	AAV20147
4	324.5	18.9	320	14	AAV3279
5	125.5	7.3	242	18	AAV2133
6	110	6.4	1657	20	AAV81355
7	110	6.4	1805	20	AAV73896
8	109	6.3	444	14	AAV34719
9	107	6.2	947	19	AAV81168
10	107	6.2	947	20	AAV07114
11	104.5	6.1	1087	20	AAV19935

12	104.5	6.1	1119	20	AAV19934
13	102.5	6.0	391	22	AAV82627
14	101.5	5.9	450	19	AAV98762
15	98.5	5.7	419	17	AAV01037
16	98	5.7	449	22	AAV80082
17	98	5.7	494	22	AAV92588
18	97.5	5.7	530	21	AAV82656
19	97.5	5.7	530	21	AAV78553
20	97.5	5.7	546	20	AAV81975
21	97.5	5.7	550	19	AAV82649
22	97.5	5.7	590	21	AAV78544
23	97.5	5.7	713	13	AAV60101
24	96.5	5.6	419	13	AAV21829
25	95.5	5.6	931	20	AAV89566
26	94	5.5	415	21	AAV09547
27	94	5.5	536	13	AAV22378
28	94	5.5	1382	20	AAV81356
29	93.5	5.4	406	6	AAV50135
30	93.5	5.4	406	18	AAV23244
31	93.5	5.4	484	8	AAV70571
32	93.5	5.4	790	19	AAV46872
33	93.5	5.4	790	20	AAV24975
34	93	5.4	282	10	AAV93505
35	93	5.4	524	20	AAV20011
36	93	5.4	553	20	AAV20010
37	93	5.4	613	21	AAV53422
38	92.5	5.4	789	19	AAV46867
39	92.5	5.4	789	19	AAV46862
40	92.5	5.4	789	20	AAV24970
41	92.5	5.4	789	20	AAV24965
42	92.5	5.4	823	19	AAV98507
43	92.5	5.4	1041	21	AAV18209
44	92	5.4	502	12	AAV15226
45	92	5.4	517	22	AAV70715

#### ALIGNMENTS

RESULT 1	AAV28460
ID	AAV28460 standard; Protein: 341 AA.
XX	AAV28460:
XX	19-OCR-1999 (first entry)
DT	P37/F1Aa protein antigen.
XX	
DE	P37; F1Aa; Lyme disease; antigen; assay; spirochete.
KW	
XX	Synthetic.
OS	Borrelia burgdorferi.
XX	
FT	Key
FT	Peptide
FT	Protein
FT	Location/Qualifiers
XX	1..22
XX	/label= signal_peptide
XX	23..341
XX	/label= mature_protein
PN	W09935272-A1.
XX	
PD	15-JUL-1999.
XX	
PF	06-JAN-1999; 99WO-US00196.
XX	
PR	08-JAN-1998; 98US-0004395.
XX	
PA	(INMR) BIOMERIEUX INC.
XX	
PI	Gilmore RD, Johnson BJ;
XX	
DR	WPI; 1999-430399/36.

B. burgdorferi ant  
S. epidermidis ope  
H. pylori GHP0 109  
Mycoplasma 46-48 k  
Corynebacterium 91  
C glutamicum prote  
Ehrlichia sp. exte  
Ehrlichia antigen  
Ehrlichia sp. W20.  
Ehrlichia sp. exte  
Canine zona pelluc  
Sequence of surfac  
Saccharomyces cere  
Arabidopsis thaila  
Antigen mc-5c. El  
A. pleuropneumonia  
Sequence of pre-pr  
Human renin. Homo  
Alpha-amylase gene  
Bacillus thuringie  
Bacillus thuringie  
Sequence encoding  
B. burgdorferi ant  
B. burgdorferi ant  
Human colon cancer  
Bacillus thuringie  
Bacillus thuringie  
Bacillus thuringie  
H. pylori GHP0 122  
Plasmid restriction  
HincII restriction  
S cerevisiae apopt

DR	N-PSDB:AAK89520.
XX	Recombinant Flaa/P37 protein antigen, useful in improved detection
PT	of Borrelia burgdorferi
XX	
PS	Claim 2; Page 26-28; 34pp; English.
XX	
CC	The sequence is the recombinant protein antigen p37/flaa. flaa is an
CC	outer sheath protein of the periplasmic flagella of borrelia burgdorferi.
CC	The p37 protein product can be used as an assay for the detection of
CC	Borrelia burgdorferi infection, the causative agent of Lyme disease. flaa
CC	is a prominent antigen in the early humoral immune response to Borrelia
CC	burgdorferi infection, and significantly suitable for use in improved
CC	serologic tests for exposure to Lyme disease spirochetes. The detection
CC	of flaa can augment the set of recombinant molecules that are recognized
CC	early in the course of disease and contribute to improved sensitivity of
CC	early testing for Lyme disease.
XX	
SQ	Sequence 341 AA:
Query Match	100.0%; Score 1719; DB 20; Length 341;
Best Local Similarity	100.0%; Pred. NO. 6.2e-162;
Matches 341; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
OY	1 MKRKASILFFLLSTVLPFAOETDGLAEGRAPGELYLDFEAELRADPSTRIDLTLTYVD 60
Db	1 mkrkakslifflstevlfagetdglaeagskraepgelvidfelaeldpsstrldltlnyvd 60
OY	61 YYSGASGIYAKEDMNVLDGINNMSVLTLPARKLAAYKANSVVAAPVAKSESKRIAGDTI 120
Db	61 yysgasgsvkpkedmndvdlgnnmsvlltpsarllqayknsvavapayvkseskryagdtl 120
OY	121 LGVRLFPESYSSGSAMIMPFKIPPYSGESGNQFLKGIDINIKTMKEITKVSYSLGPER 180
Db	121 lgvrlfpesygssasamimpfkipfysgesgnqlfgklidniktmkelkvsylsygei 180
OY	181 DLEVLPEDMNGMEIVYSMGTLFKGMADLI MSNPNTIPINISIRIIKDVPNPPLASSKMR 240
Db	181 dlevlpedmngmeivaysmgtlfkfwadliwspnyipnlissriliikddvpnpplasskmr 240
OY	241 FAFARSRKSHSKVNKFIPYVKDLRVLYDKLSVISDISDSISVFKEYPTSGETSLRKLK 300
Db	241 fafarvskshskvnkfipyvkdlrvlydklsvisdisdsisvfkeyptsgetslrklk 300
OY	301 AHETERKVLKIREKTISIAGSFONFEVETIESKPEESSPKN 341
Db	301 aheterkvlikrekisiagsfqnfevekiesekpeesphn 341
RESULT 2	
AAV20146	
ID AAV20146	standard; Protein; 335 AA.
AC AAY20146;	
XX	
DT 19-JUL-1999	(first entry)
XX	
DE B. burgdorferi antigenic protein, f07A.a	BB023.
KW Antigenic protein; vaccine; Lyme disease; infection; detection.	
XX	
OS Borrelia burgdorferi.	
XX	
PN W09B59071-A1.	
PD 30-DEC-1998.	
XX	
PF 18-JUN-1998;	98WO-US12718.
XX	
PR 03-SEP-1997;	97US-0057483.
PR 20-JUN-1997;	97US-0050359.
PR 22-JUN-1997;	97US-0053344.

Query	Match	Best Local Similarity	95.7%: Score 1645; DB 20: Length 335;
Matches 327; Conservative	2; Mismatches 6; Indels 0; Gaps 0		
7 SLIFPLSLTVLEAQTGDAIEGSKRAEPGLVLDFAELARDPSSTRLLDTNYYVYVYSGA 66	1 SLIFPLSLTVLEAQTGDAIEGSKRAEPGLVLDFAELARDPSSTRLLDTNYYVYVYSGA 60		
67 SGIVRPEDMDVVDLGINNMSVLLTPSARLQAVYKNSVYAPAVYKSESKRYAGDRIIGVRVL 126	61 SGIVRPEDMDVVDLGINNMSVLLTPSARLQAVYKNSVYAPAVYKSESKRYAGDRIIGVRVL 120		
127 PFSYSSOSSAMIPPEKIPRYSGESGNOFGKGLIDNITKMKIEIKVYSLSIGYEIDLEVL 186	121 PFSYSSOSSAMIPPEKIPRYSGESGNOFGKGLIDNITKMKIEIKVYSLSIGYEIDLEVL 180		
187 EMMNGMEYAVSMGTLKFKGMADLWNSNPYIRNISRIITKDVPNPLPSSKMRKFAFRV 246	181 EMMNGMEYAVSMGTLKFKGMADLWNSNPYIRNISRIITKDVPNPLPSSKMRKFAFRV 240		
247 SKSHSKVYVNFYFYKDLVLDKLSVSDSIDSESVFKVETSGTESLRLKLAHEFEK 306	241 SKSHSKVYVNFYFYKDLVLDKLSVSDSIDSESVFKVETSGTESLRLKLAHEFEK 300		
307 RVLKLRKISIAEGSFONFVEKIESKPESSPKN 341	301 RVLKLRKISIAEGSFONFVEKIESKPESSPKN 335		
AAV20147	AAV20147 standard; Protein; 315 AA.		
AAV20147;	19-JUL-1999 (first entry)		
B. burgdorferi antigenic protein, t07A.aa B8023.	Antigenic protein; vaccine; Lyme disease; Infection; detection.		
Borrelia burgdorferi.			
W09859071-A1.			
30-DEC-1998.			



XX 18-JUN-1998; 98WO-US12718.  
XX  
XX 03-SEP-1997; 97US-0057483.  
PR 20-JUN-1997; 97US-0050359.  
PR 22-JUL-1997; 97US-0053444.  
PR 22-JUL-1997; 97US-0053377.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (MEDI-) MEDIMUNE INC.  
XX  
XX Chai GH, Erwin AL, Hanson MS, Lathigra R;  
XX  
XX MPI: 1999-189980/16.  
DR N-PSDB; AAK61844.  
XX  
XX New isolated Borrelia burgdorferi nucleic acids - used to develop  
PT products for the diagnosis, prevention and treatment of diseases  
PT caused by Borrelia, particularly Lyme disease  
XX  
XX Claim 12; Page 214; 275pp; English.  
XX  
XX This sequence represents a Borrelia burgdorferi (Bb) protein of the  
CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
CC can be used in vaccines for eliciting protective antibodies to members of  
CC the Borrelia genus, particularly for the use against Lyme disease in  
CC humans and animals. They can be used for preventing or attenuating an  
CC infection caused by a member of the Borrelia genus. The products can also  
CC be used for detection of members of the Borrelia genus.  
XX  
SQ Sequence 315 AA;

Query Match 90.2%; Score 1551; DB 20; Length 315;  
Best Local Similarity 97.5%; Pred. No. 2.5e-145;  
Matches 307; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
QY 27 ESKRAPEPELVDFAEELARDPSSTRDLDTNVDYVYSGASGIVKEDMVDLGINMSV 86  
DB 1 ESKRAPEPELVDFAEELARDPSSTRDLDTNVDYVYSGASGIVKEDMVDLGINMSV 60  
QY 87 LTPSARLQAVYKNSVAVPAVVKSESKRYAGDTILGVRVLPSPYSOSSAMIMPFKIPRY 146  
DB 61 ILTPSARLQAVYKNSVAVPAVVKSESKRYAGDTILGVRVLPSPYSOSSAMIMPFKIPRY 120  
QY 147 SEESGNQFLGKGLIDNKTKEIKVSVSLGYEIDLEVLFEEDMNGMEYAYSMGTLEKFGW 206  
DB 121 sgesgnqflgkglidnktkeikvsvslgyeidlevlfednmngmeyaxsmgtlkfkgw 180  
QY 207 ADLINSNPYIRINISRIITKDDVPNPLASSKMKRKAFFVSKSHSKVNFTFYVDLKV 266  
DB 181 adlinsnpyirinisriitkddvpnplasskmmrkafvskshskenflftfyvkdllv 240  
QY 267 LYDKLSVSDSIDSESVKVEYETSGTESLRKLKAHETFKRVLKLRKISIAESQNFY 326  
DB 241 lydklsvdsidsdesvkvkyetsgteslrklkxhtkxvklkrekismpegsfqnfy 300  
QY 327 EKIESKPEESSPKN 341  
DB 301 ekiesekpeespskn 315  
RESULT 4  
ID AAR33279 standard; Protein; 320 AA.  
XX  
XX AAR33279;  
AC  
XX  
XX 16-JUL-1993 (first entry)  
DT  
XX  
XX 43 kD endoflagellum sheath protein.  
DE  
XX  
XX Endoflagellum; sheath protein; T. hyodysenteriae; core; antibody;

KW bacteriacide; 43 kD; vaccine; infection; swine dysentery.  
XX  
XX Treponema hyodysenteriae.  
OS  
XX  
XX Key Location/Qualifiers  
FH Peptide 1..19  
FT /note="Signal peptide"  
FT Protein 20..320  
FT /note="Mature protein"  
XX  
XX EP534526-A.  
PN  
XX  
XX 31-MAR-1993.  
PD  
XX  
XX 14-SEP-1992; 92EP-0202796.  
PF  
XX  
XX 25-SEP-1991; 91EP-0202478.  
PR 24-JUL-1992; 92EP-0202273.  
PR  
XX  
XX (DUTN ) DUPHAR INT RES BV.  
PA  
XX  
XX Koopman MBH, Kusters JG;  
PI  
XX  
XX MPI: 1993-102665/13.  
DR N-PSDB; AAK38583.  
XX  
XX Vaccine to protect pigs against swine dysentery - comprises  
PT Treponema hyodysenteriae endo-flagellum sheath protein, applied  
PT orally or intranasally  
XX  
XX  
XX Claim 2; Page 21-22; 34pp; English.  
XX  
XX This sequence represents the endoflagellum sheath protein of T.  
CC hyodysenteriae. The endoflagellum consists of at least four  
CC proteins, this protein forms the sheath of the flagellum and three  
CC proteins, of molecular weights 37, 34 and 32 kD, make up its core.  
CC Antibodies raised against the sheath protein have been shown to be  
CC bactericidal for T. hyodysenteriae. The 43 kD sheath protein can  
CC be used in the production of a vaccine against infections such as  
CC swine dysentery.  
XX  
SQ Sequence 320 AA;

Query Match 18.9%; Score 324.5; DB 14; Length 320;  
Best Local Similarity 28.4%; Pred. No. 1e-23;  
Matches 89; Conservative 63; Mismatches 130; Indels 31; Gaps 8;  
QY 9 LPEFLSTVLFAPQETDGLAESKRAPEPELVDFAEELARDPSSTRDLDTNVDYVYSGASG 68  
DB 4 LTVLSTVLFAPQETDGLAESKRAPEPELVDFAEELARDPSSTRDLDTNVDYVYSGASG 43  
QY 69 IYKPEDMVVDLGINMSVLLTPSARLQAVYKNSVAVPAVVKSESKRYAGDTILGVRVLP 128  
DB 44 thevypvaenlyndvrvvlnesarltenrnsvynvdskgngawagkvlgvrvhfp 103  
QY 129 SYSQSS-AMIMPFKIPFTSGESGNQFL-GKGLIDNKTKEIKVSVSLGYEIDLEVL 186  
DB 104 laamsyalyvktvylemgygagdktyteggkgyidhvgelksisawvygrnyllsyfvnl 163  
QY 187 EDMNGMEYAVSMGTLEKFGKADLIMSNPNYIPNISRRIITKDDVPNPLASSKMKRKAFFRV 246  
DB 164 qnefgelksypmgltvyfngyrgvwenreylnpvrdsavlre-plyprmlpsvkidslgtf 222  
QY 247 SKSHSKVKNFIYVVDLRYLDKLSVSDSIDSESVKVEYETSGTESLRKLKAHETEK 306  
DB 223 yrtkdktgsgdfltyvkdvtleydvvvvvdfeediddeatwqjlk---tendrk-gates-- 276  
QY 307 RVLKLREKISIAE 319  
DB 277 --arlrgealrd 287

```

RESULT 5
AAW22133
ID AAW22133 standard; Protein: 242 AA.
XX
XX AAW22133;
AC
XX 12-JAN-1998 (first entry)
DT
XX Treponema pallidum rare outer membrane protein (TROMP-2).
DE
XX Rare: outer membrane protein; Treponema pallidum; Spirochaetales;
KM Immune response; Syphilis; TROMP-2; precursor.
XX
XX Treponema pallidum.
OS
XX Key
FH Location/Qualifiers
FT Peptide 1..24
FT /note= "putative signal peptide"
FT Cleavage-site 22..24
FT /note= "cleaves at the C-terminal side of Ala 24"
FT Protein 25..242
FT /note= "putative mature protein"
XX
XX W09727215-A1.
PN
XX 31-JUL-1997.
PD
XX 23-JAN-1997; 97WO-US01302.
PE
XX 23-JAN-1996; 96US-0599480.
PR
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
PA (REGC ) UNIV CALIFORNIA.
XX
XX Blanco DR, Lovitt MA, Miller JN, Tempst PJ;
PI WPI: 1997-393614/36.
XX N-PSDB; AAT76579.
DR
XX
XX New isolated Treponema pallidum outer membrane protein - used in the
PT detection and induction of immune response in an animal to
PT pathogenic Spirochaetales
XX
XX Claim 1; Fig 11; 77pp; English.
PS
XX This rare outer membrane protein (OMP) of species Treponema pallidum,
CC subspecies pallidum is encoded by a 765 bp sequence (TROMP2 gene).
CC This is a precursor TROMP2 protein of 242 amino acids. Treponemes
CC were extracted from infected rabbit testicles and the suspension
CC tissue debris. Treponemes were obtained from the uppermost zone
CC gradient after two high speed centrifugations involving a density
CC gradient separation medium. R18 octyl-decyl Rhodamine chloride was added
CC to the treponemal suspension which was incubated at room temperature for
CC 10 minutes and then centrifuged. The treponemal pellet was resuspended
CC in ice cold sodium citrate buffer, incubated with occasional
CC vortexing to release the outer membrane from the inner membrane
CC Treponemal protoplasmic cylinders were removed by centrifugation.
CC The outer membrane in the supernatant was subjected to SDS-PAGE to
CC obtain the rare OMP. The recombinant expression of the rare OMP's
CC can be used for diagnostic tests to detect syphilis and for
CC development of host immunity during syphilis.
XX
XX Sequence 242 AA;
SO

```

Query Match 7.3%; Score 125.5; DB 18; Length 242;  
 Best Local Similarity 22.4%; Pred. No. 0.00036;  
 Matches 55; Conservative 31; Mismatches 104; Indels 55; Gaps 8;

```

QY 48 PSSSTRLDITNTVDYV---YSGASGIVKPEDWVDIGINN---SVLTTPSARLQAVYKN 100
DB 21 plaagrskvnygayfiddfdgas-----edggl-awraagskftkqfpllkfyeg 70

```

```

QY 101 SVAPAVVKSSEKRYAGDTILGVEVLEPPSYSQSSAMIMP-----PKRIPEYSGESGNOFL 155
DB 71 mpqavrmagsgwqgkdkeariflyvecklnrgqnmwldlptkqgsdyelpirgvsg----- 126
QY 156 GKGLIDNITKWEIKVSYSLGYEIDLEVLFPEDMNGMEYATSMGTLFKMGADLIRNSPN 215
DB 127 -----fdvwwagqyyslaalvrdctgrvthlllgldfgqwknlsvsupt 173
QY 216 YIPNISRRIKDVDPNYPPLASSKMRFAFRVSKSHSSKRVKRFYVDLRYLXKLSYSI 275
DB 174 htpqts-----rylgsqhlstfygfrlrrtspservdffy-----vffqgfakla 217
QY 276 DSDID 280
DB 218 nmhid 222

```

```

RESULT 6
AAW81355
ID AAW81355 standard; Protein: 1657 AA.
XX
XX AAW81355;
AC
XX 16-APR-1999 (first entry)
DT
XX A. pleuropneumoniae ApxIV toxin serotype 1 protein sequence.
DE
XX
XX ApxIV toxin; RTX toxin; attenuated microorganism; subunit vaccine;
KW A. pleuropneumoniae infection; diagnosis.
XX
XX Actinobacillus pleuropneumoniae.
OS
XX EP875574-A2.
PN
XX 04-NOV-1998.
PD
XX 08-APR-1998; 98EP-0201115.
PE
XX 10-APR-1997; 97EP-0201032.
PR
XX (ALKU ) AKZO NOBEL NV.
PA
XX Frey J, Segers RPAM;
PI
XX WPI: 1999-027956/03.
DR N-PSDB; AAV68471.
XX
XX Attenuated live Actinobacillus pleuropneumoniae - and a nucleotide
PT sequence having promoter controlling expression of ApxIV gene, a
PT sub-unit vaccine containing pure ApxIV toxin, and an attenuated live
PT vaccine
XX
XX Example 1; Page 13-20; 56pp; English.
PS
XX This sequence represents the ApxIV toxin serotype 1 isolated from
CC Actinobacillus pleuropneumoniae strain 4074 (serotype 1 reference
CC strain). The invention relates to an attenuated live Actinobacillus
CC pleuropneumoniae microorganism producing no functional ApxIV toxin. The
CC invention also provides a subunit vaccine for protecting an animal
CC against infection of A. pleuropneumoniae, the vaccine contains pure ApxIV
CC toxin and a pharmacologically allowable carrier. The pure ApxIV toxin can
CC be used for the diagnosis based on the presence of an antibody against
CC ApxIV for distinguishing A. pleuropneumoniae infection from A. suis.
CC N.B. This sequence was indexed from EP-875574, which is the first major
CC country equivalent to JP10290695.
XX
XX Sequence 1657 AA;
SO

```

Query Match 6.4%; Score 110; DB 20; Length 1657;  
 Best Local Similarity 21.5%; Pred. No. 0.25;  
 Matches 68; Conservative 51; Mismatches 114; Indels 84; Gaps 17;



CC (MCS). Analysis of the sequence showed four regions potentially  
CC coding for proteins, a zone upstream of the first ORF contg. the  
CC sRNA operon promoter and a presumed terminator positioned downstream  
CC of the stop codon of the fourth ORF. Chromosomal DNA downstream of  
CC the ORF4 stop codon was also sequenced, sharing four regions (ORF5-8)  
CC potentially coding for different prods. ORF6 and ORF7 are contiguous  
CC and are potentially transcribed in the opposite direction to sRNA from  
CC a single promoter situated between ORF7 and ORF8. ORF8 is  
CC transcribed in the same direction as the sRNA operon. A further  
CC two potential coding sequences were identified in the region immediately  
CC downstream of ORF4, namely ORFX and ORFY.  
CC See also AAR34712-21.

CC Sequence 444 AA;

Query Match 6.3%; Score 109; DB 14; Length 444;  
Best Local Similarity 21.6%; Pred. No. 0.041;  
Matches 75; Conservative 42; Mismatches 126; Indels 104; Gaps 15;

OY 29 SKRAPGELVDFEALARDPSSRLDLNLYVD-----YVSGASGI--- 69  
DB 73 sksgqpgp--idfatsapdp-----dvfpjldfphcinkaidtykndifiytpkylpsl 125  
OY 70 -----VKPEDMVVDLGIMNMSVLL-----TPSARLQAVYKN 100  
DB 126 lrvllkllatqgvfaderhifltsvgqalsllcampfnqekialegpyhlwveqe 185  
OY 101 SYVAPAVVKSSEKRYAGDTILGVRVLPSPYSOSSAMIMPFKIPYSGESGNQFLGKGI 160  
DB 186 tlgjpaigvkrtee--gldiakverllfqesikffiytmprfmp-----lgscls 233  
OY 161 DNIKMKEIKVSVSYSLGYEIDLEVFEDMNGMEYAVSMGTLFKGMDLIMSNPVTIPNI 220  
DB 234 eddkq-----eivrlaaydyvivedy-----lgldeenkkad-----plyaydl 274  
OY 221 SRIIKDDVPNYPLASSKMKRFKAFVSKSHSKVKNFIIFYVDLRYLDKLSVSDSD 280  
DB 275 sashvl-----ylksfsmkmpfplrgaavlpcaaltftya-----ykhln-dldcsm 321  
OY 281 SSVSVKRYETSGTESLRLKAHETFK-RYLKIREKI-----SIAEGSF 322  
DB 322 sqaalelylksqmygrhkekrldsykerslrhgaalrthrglsgrf 368

RESULT 9

AAW81168 standard; Protein: 947 AA.

AAW81168;

05-MAR-1999 (first entry)

Transcriptional regulatory factor RING3.

Human; transcriptional regulatory factor; RING3; TSB; cancer;  
testis specific bromodomain; testicular cell proliferation.

Homo sapiens.

MO9848015-A1.

29-OCT-1998.

17-APR-1998; 98WO-JP01782.

18-APR-1997; 97JP-0116402.

(CHUGAI) CHUGAI RES INST MOLECULAR MEDICINE INC.

Jones MH;

WPI; 1998-583658/49.

DR N-PSDB; AAV68343.

Transcriptional regulator gene containing bromodomain sequence - may  
be expressed in testis tissue and is useful in treatment of cancer  
and other proliferative disorders

Claim 1; Page 19-24; 42pp; Japanese.

The present sequence represents the human transcriptional regulatory  
factor RING3, which is isolated from testicular cells. RING3 contains  
a testis specific bromodomain (TSB) which is expressed specifically  
in testis tissue and also expressed in certain tumour lines. The  
transgenic cells may be used to express RING3 which is a TSB expression  
protein. The TSB expression product can be used in the treatment of  
cancer and other proliferative disorders, and in screening of compounds  
for ability to bind to it (e.g. for use as drugs by modulation of  
transcriptional regulation). DNA capable of hybridising to RING3  
polynucleotides may be used for construction of probes and primers.

Sequence 947 AA;

Query Match 6.2%; Score 107; DB 19; Length 947;  
Best Local Similarity 18.5%; Pred. No. 0.21;  
Matches 76; Conservative 80; Mismatches 120; Indels 134; Gaps 20;

OY 1 MKRKASILFPLSTVFLVAOENHGLAEGSKRAPEGLVDFEALARDPSSR-RLDLTNYV 59  
DB 118 maqalekifmgklsqmpgeevgyvkerikytqgnlvsasakesspsalekvtkgel 177  
OY 60 DYVVGASGIVKPEDMVVDLGIMNMSVLLTPSARLQAVYK-----NSVAPAVVKSSEKRY 115  
DB 178 psvfkte--lspnlvvgasvsnss--qtaagvtkgvkrkadtltpatsavkase-- 230  
OY 116 AGDTILGVRVLPSPYSOSSAMIMPFK-----IPYSGESGNQFLGKGLIDNKTKE 168  
DB 231 -----fslptltek--valpikemmpknvlp-----dsqgy-----nvatkvteq 272  
OY 169 IK-----VSYSLG-----YEI-----DLVLFEDMNGMEY- 194  
DB 273 lthceilkemlakhhfyaepfynpvdnalglnydvknnpndlytlkekndngyk 332  
OY 195 -AYSMTLKFKGWADLIMSN-----PNYIPNISRIIKDDVPNYPLASSKMKRFKAFV 247  
DB 333 days-----faadvrlmfmncykynpdphevvtumarmld-----vf 369  
OY 248 KSHSKVKNFIIFYVDLRYLDKLSVSDSDSDSVKRYETSGTESLRLKAH----- 302  
DB 370 ethfskip--lepyesmplcyikldt-----etlgrentnaesegnssd 413  
OY 303 ----ETFKRVILKREKIS-----IAEGSPONFEKIESKPESSPK 340  
DB 414 deederavkrlaklqeglkvavngqlvlsqvfpfirklnkkkkskkkkkxk 463

RESULT 10

AAW07114 standard; Protein: 947 AA.

AAW07114;

02-JUL-1999 (first entry)

WO9904265 Seq ID No: 685.

Cancer associated antigen; diagnosis; research; treatment; human;  
breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
prostate cancer.

Homo sapiens.

MO9904265-A2.

PD 28-JAN-1999.  
XX  
PF 15-JUL-1998; 98WO-US14679.  
XX  
PR 22-JUN-1998; 98US-0102322.  
PR 17-JUL-1997; 97US-0896164.  
PR 10-OCT-1997; 97US-0061599.  
PR 10-OCT-1997; 97US-0061765.  
PR 10-OCT-1997; 97US-0948705.  
PR 11-OCT-1997; 97GB-0021697.  
XX  
PA (LUDW-) LUDWIG INST CANCER RES.  
XX  
PI Chen Y, Gout I, Gure A, O'Hare M, Opatka Y, Old LJ;  
PI Pfeundschnuh M, Sahlin U, Scanlan MJ, Stockert E;  
PI Tureci O;  
XX  
DR WPI: 1999-132448/11.  
XX  
PT New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
XX  
PS Disclosure: Page 728-730; 787pp; English.  
XX  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
XX  
SQ Sequence 947 AA:

Query Match 6.2%; Score 107; DB 20; Length 947;  
Best Local Similarity 18.5%; Pred. No. 0.21;  
Matches 76; Conservative 80; Mismatches 120; Indels 134; Gaps 20;

1 MKRKAASILFELSTVLFPAQETDGLAEGSKRAPEGLVDFEALARPESST-RLDLTNYV 59  
118 maqaleklfmqqlsqmpgeevgyvkerlkgltqgnlavsakeksspatekvfkqgei 177  
DB  
QY 60 DYVVGASGIVPEDMVVDLGINNMSVLLTPSARLQAVYK---NSVAPAVYKSSKRY 115  
DB 178 psvfptks--lspilnvvgasvss--qtaaytkykrkadtltptsavkaese-- 230  
QY 116 ADDTLIGVAVLFPSYSQSSAMIMPEK-----IPFYSGEGNOFLGGLDNTITMKE 168  
DB 231 -----fspitltxs-valpikempxnvlp---dsqgqy--nvvetvvtqg 272  
QY 169 IK-----VSYSLG---YEI-----DLEVLFEPMNGMEY- 194  
DB 273 lthcseellkemlakkhfsyapfynpvdvnaiglmydvknpmldglclikekmdqgk 332  
QY 195 -AYSMTLKFQKQWADLIMSN-----PNYIPNISRIKDDVPNYPPLASSKMKRFAFRYS 247  
DB 333 days-----faadvrlmfmycykynppdhevvtmarlqg-----Vf 369  
QY 248 KHSKSKVKNFIRYVVDLRLVLDKLSVSDSDSESVFVYETSGTESLRKLKAH----- 302  
DB 370 ethfsklip--lepvasmplyikltdt-----elttgenennessesgnsad 413  
QY 303 -----ETFKRVLTLRKIS-----IAGSFONFVEKIESEKPESSPK 340

DB 414 dsedevrklaklgqllkavhqlqylsqvprfklnkkkkekkekkek 463

RESULT 11  
AA19935  
ID AA19935 standard; Protein; 1087 AA.  
XX  
AC AA19935;  
XX  
DT 19-JUL-1999 (first entry)  
XX  
DE B. burgdorferi antigenic protein, t742.aa.  
XX  
KM Antigenic protein; vaccine; Lyme disease; Infection; detection.  
XX  
OS Borrelia burgdorferi.  
XX  
PN W09859071-A1.  
XX  
PD 30-DEC-1998.  
XX  
PF 18-JUN-1998; 98WO-US12718.  
XX  
PR 03-SEP-1997; 97US-0057483.  
PR 20-JUN-1997; 97US-0050359.  
PR 22-JUL-1997; 97US-0053344.  
PR 22-JUL-1997; 97US-0053377.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI (MEDI-) MEDIMUNE INC.  
XX  
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;  
XX  
DR WPI: 1999-189980/16.  
XX  
DR N-PSDB; AAX61632.  
XX  
PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
PT products for the diagnosis, prevention and treatment of diseases  
PT caused by Borrelia, particularly Lyme disease  
XX  
PS Claim 12; Page 125; 275pp; English.  
XX  
CC This sequence represents a Borrelia burgdorferi (Bb) protein of the  
CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
CC can be used in vaccines for eliciting protective antibodies to members of  
CC the Borrelia genus, particularly for the use against Lyme disease in  
CC humans and animals. They can be used for preventing or attenuating an  
CC infection caused by a member of the Borrelia genus. The products can also  
CC be used for detection of members of the Borrelia genus.  
XX  
SQ Sequence 1087 AA:

Query Match 6.1%; Score 104.5; DB 20; Length 1087;  
Best Local Similarity 21.7%; Pred. No. 0.46;  
Matches 68; Conservative 44; Mismatches 110; Indels 91; Gaps 12;

98 VKNSVAPAVVYKSSKRYAGDTILGAVLFPSYSQSSAMIM-----PPFKRIPY 146  
DB 14 vknsvid-----nykkays-----vakllqdkypqedlamlntlaeanspfeeskd 64  
QY 147 SGESGQFLG--KGLDNIKT-----MKELKVSYSIGY-EIDLEVLFEDMNGM 192  
DB 65 grdsaqllidklkqg-dnkttnvenfdlafmrylkdsltenysdrnddvyiededs 123  
QY 193 EYAYSMTLKFQKQWADLIMSNPNYIPNISRIKDDVPNYPPLASSKMKRFAFRYSKSHS 252  
DB 124 ethfsklip-----pntmpkeedqllsqpnklsvndqknlfnlekikknlsq 174  
QY 253 KYKNFIFYVVDLRLVLDKLSVSDSDSESVFVYETSGTESLRKL----- 299  
DB 175 kansenllndsqllendkqntlsksknsenllktpdnakysnmmttslkkksnsgke 234

```

Oy 300 -----KAHEFEKRRVLT-----LRKRISIA-EGSQONE 325
      | : : : |
Db 225 selpspsttltgkryrpsytlkelyellddintgrvtlgnlkelikkjglsnktqkv 294
      | : : |
Oy 326 VEKIESEKPEEES 338
      | | : |
Db 295 nellenskneas 307

```

```

RESULT 12
AAV19934
ID AAV19934 standard; Protein; 1119 AA.
XX
AC AAV19934;

```

DT 19-JUL-1999 (first entry)  
XX  
DE B. burgdorferi antigenic protein, f742.aa.  
XX  
KW Antigenic protein; vaccine; Lyme disease; infection; detection  
XX  
XS Borrelia burgdorferi.

PN	W09859071-AL	
XX		
PD	30-DEC-1998.	
XX		
PF	18-JUN-1998;	98WO-US12718.
XX		
PR	03-SEP-1997;	97US-0057483.
PR	20-JUN-1997;	97US-0050359.
PR	22-JUL-1997;	97US-0053344.
ER	22-JUL-1997;	97US-0053377.

PA (HUMA-) HUMAN GENOME SCI INC.  
PA (MEDI-) MEDIMMUNE INC.

PI Choi GH, Erwin AL, Hanson MS, Lathigra R;

DR WPI; 1999-189980/16.  
DR N-PSDB; AAX61631.

PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop  
PT products for the diagnosis, prevention and treatment of diseases  
PT caused by *Borrelia*, particularly Lyme disease

PS Claim 12; Page 124-125; 275pp; English

CC This sequence represents a *Borrelia burgdorferi* (Bb) protein of the  
CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
CC can be used in vaccines for eliciting protective antibodies to members of  
CC the *Borrelia* genus, particularly for the use against Lyme disease in  
CC humans and animals. They can be used for preventing or attenuating an  
CC infection caused by a member of the *Borrelia* genus. The products can also  
CC be used for detection of members of the *Borrelia* genus.

SQ Sequence 1119 AA;

Query Match	6.1%;	Score 104.5;	DB 20;	Length 1119;
Best Local Similarity	21.7%;	Pred. No. 0.48;		
Matches 68;	Conservative 44;	Mismatches 110;	Indels 91;	Gaps 12

```

QY      98 VKNSVAPAVKSESKRYACDTILGVRVLFPSYSOSSAMIM-----PPKIPFY 146
      ||||:  :: | ||:  ::|| | ||:  :: | ||:

```

QY 147 SGESENPLG--KGLIDNIKT-----MKELKSVSYSLGY-EDLEVLFEEDNGM 1922  
: || | || || : || | : : : || :

QY 193 EYAYSMGTLKFKGNADLIWSNPNIPIPTISSRIIKDDVPNYPYLASSKMKREKAFRVSKSHS 252Z

Db	156	effkskskipek1x-----	putnpxkeedq1iqspmkxisvndqkmlfnlcklknlsq	206
Oy	253	KVKNFIEYVQDLRVLYDVKLSVSDSDSDSDSVKRVYETS-----	GTESIRKL-----	299
Db	207	knsenl1ndsqldqgnfnlsskexnsenl1ktpdskysnmnnltskissnqke		266
Oy	300	-----KAHERFKRRLK-----	LREKISIA-EGSPQNF	325
Db	267	selsppsqtl1gk1ytrpysyl1kkel1yellodd1ntgrv1lgkml1kellk1g1ank1qkv		326
Oy	326	VEKTESSEKPESS 338		
Db	327	nellensknkeas. 339		

RESULT	13
AAG82627	
ID	AAG82627 standard; Protein; 391 AA

DT 03-SEP-2001 (first entry)

DE	S.	epidermidis	open	reading	frame	protein	sequence	SEQ ID NO:	2348.
----	----	-------------	------	---------	-------	---------	----------	------------	-------

KW Staphylococcus epidermidis SRI strain; infection; diagnosis,  
KW vaccination; endocarditis.

OS Staphylococcus epidermidis.

PN W0200134809-A2

PD 17-MAY-2001.

PF 09-NOV-2000; 2000WO-US30782.

PR 09-NOV-1999; 99US-0164258.

PA (GLAX ) GLAXO GROUP LTD.

PI Kimmerly WJ;

DR WPI; 2001-316495/33

XX  
33

PT useful for vaccinating against infections, e.g. endocarditis -

PS Claim 18; Page 621; 2188pp; English.

AAH5230 and AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AG681454 to AG681320, from *Staphylococcus epidermidis*. CC  
CC (I) and (II) can have antihistacterial activity and therefore can be used CC  
CC in vaccination. The nucleic acids (I) may be used to produce the CC  
CC S. epidermidis polypeptides (II) via the production of vectors CC  
CC containing them which are used to produce hosts cells which express the CC  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be CC  
CC used to vaccinate subjects and to raise antibodies against the bacteria CC  
CC. The polypeptides may also be used to assay for other inhibitors of their CC  
CC activity and therefore identify compounds that may be used for the CC  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to CC  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA CC  
CC polynucleotide sequences from the present invention. AAH55091 to CC  
CC AAH55098 represent oligonucleotide sequences and primers which are used CC  
CC in the exemplification of the present invention. CC  
CC N.B. The present invention specifically claims all the polynucleotide CC  
CC sequences given in the sequence listing of the present specification. CC  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even CC  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, CC  
CC no sequences are present for SEQ ID NO:4455 to 4464. CC

SQ Sequence 391 AA;







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2002, 00:00:39 ; Search time 1691.33 Seconds  
(Without alignments) 842.965 Million cell updates/sec

Title: US-09-004-395-1

Perfect score: 1663  
Sequence: 1 atgataacttttttcaaaa.....atatacttgaagagaataatt 1663

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: N\_Geneseq\_1101.\*  
2: /SIDS2/gcgcdata/geneseq/geneseq/NA1980.DAT.\*  
3: /SIDS2/gcgcdata/geneseq/geneseq/NA1981.DAT.\*  
4: /SIDS2/gcgcdata/geneseq/geneseq/NA1982.DAT.\*  
5: /SIDS2/gcgcdata/geneseq/geneseq/NA1983.DAT.\*  
6: /SIDS2/gcgcdata/geneseq/geneseq/NA1984.DAT.\*  
7: /SIDS2/gcgcdata/geneseq/geneseq/NA1985.DAT.\*  
8: /SIDS2/gcgcdata/geneseq/geneseq/NA1986.DAT.\*  
9: /SIDS2/gcgcdata/geneseq/geneseq/NA1987.DAT.\*  
10: /SIDS2/gcgcdata/geneseq/geneseq/NA1988.DAT.\*  
11: /SIDS2/gcgcdata/geneseq/geneseq/NA1989.DAT.\*  
12: /SIDS2/gcgcdata/geneseq/geneseq/NA1990.DAT.\*  
13: /SIDS2/gcgcdata/geneseq/geneseq/NA1991.DAT.\*  
14: /SIDS2/gcgcdata/geneseq/geneseq/NA1992.DAT.\*  
15: /SIDS2/gcgcdata/geneseq/geneseq/NA1993.DAT.\*  
16: /SIDS2/gcgcdata/geneseq/geneseq/NA1994.DAT.\*  
17: /SIDS2/gcgcdata/geneseq/geneseq/NA1995.DAT.\*  
18: /SIDS2/gcgcdata/geneseq/geneseq/NA1996.DAT.\*  
19: /SIDS2/gcgcdata/geneseq/geneseq/NA1997.DAT.\*  
20: /SIDS2/gcgcdata/geneseq/geneseq/NA1998.DAT.\*  
21: /SIDS2/gcgcdata/geneseq/geneseq/NA1999.DAT.\*  
22: /SIDS2/gcgcdata/geneseq/geneseq/NA2000.DAT.\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1637	98.4	1655	20	AAx89520
2	989.6	59.5	1011	20	AAx61843
3	924.6	55.6	945	20	AAx61844
4	372.8	22.4	423	20	AAx61740
5	372.8	22.4	474	20	AAx61739
6	207	12.4	936	22	AAx58252
7	207	12.4	936	22	AAx58254
8	207	12.4	936	22	AAx58257
9	207	12.4	936	22	AAx58259
10	207	12.4	936	22	AAx58262
11	207	12.4	938	22	AAx58255

C	12	206	12.4	936	22	AAx58252	Oligonucleotide D1
C	13	206	12.4	936	22	AAx58254	Oligonucleotide D1
C	14	206	12.4	936	22	AAx58257	Oligonucleotide D1
C	15	206	12.4	936	22	AAx58259	Oligonucleotide D2
C	16	206	12.4	936	22	AAx58262	Oligonucleotide D2
C	17	206	12.4	938	22	AAx58255	Oligonucleotide D1
C	18	107	6.4	1140	14	AAO38583	43 kd endoflagellin
C	19	78	4.7	19124	18	AAx72882	Plasmodium var-7 g
C	20	78	4.7	19124	21	AAx82827	Plasmodium var-7 p
C	21	75.6	4.5	244	22	AAx58238	Oligonucleotide D1
C	22	72.6	4.4	244	22	AAx58238	Oligonucleotide D1
C	23	71.4	4.3	612	22	AAx71471	Human cervical can
C	24	67	4.0	6124	11	AAO03568	Sequence encoding
C	25	64.8	3.9	3975	9	AAx81157	Malaria-specific g
C	26	64.8	3.9	3975	13	AAQ22999	SCR gene, Plasm
C	27	63.6	3.8	723	22	AAx71575	Human cervical can
C	28	63	3.8	665	21	AAx231996	Human apolipoprote
C	29	61.6	3.7	335913	22	AAx61371	Soybean 240017 reg
C	30	61.6	3.7	335913	22	AAx61372	Soybean 240017 reg
C	31	60.4	3.6	3567	7	AAx70117	Plasmodium falcipa
C	32	58.8	3.5	1444	21	AAx24422	Plasmodium falcipa
C	33	58.4	3.5	4590	7	AAx60472	Sequence encoding
C	34	58.2	3.5	4056	21	AAx70225	Plasmodium falcipa
C	35	58.2	3.5	10640	22	AAO3729	Sequence encoding
C	36	57.8	3.5	1612	7	AAx60392	Sequence encoding
C	37	57.8	3.5	2418	13	AAQ27886	P. falciparum GBP13
C	38	57.6	3.5	2104	13	AAQ25273	Sequence encoding
C	39	57.6	3.5	5454	21	AAx70236	Plasmodium falcipa
C	40	57.2	3.4	9789	17	AAx71852	cDNA encoding plas
C	41	56.8	3.4	605	17	AAx71530	Human 3' apolipopr
C	42	56.8	3.4	6124	11	AAO03568	Sequence encoding
C	43	56.2	3.4	513445	22	AAx61373	Soybean 318013 reg
C	44	56	3.4	1251	21	AAx70119	Plasmodium falcipa
C	45	56	3.4	8920	15	AAO62924	Carbamoyl-phosphat

## ALIGNMENTS

RESULT 1	AAx89520
ID	AAx89520 standard; DNA; 1655 BP.
AC	AAx89520;
XX	19-OCR-1999 (first entry)
DT	
XX	Nucleic acid sequence encoding the Flaa gene product.
DE	
XX	P37; Flaa; Lyme disease; antigen; assay; splirochete; ss.
XX	Synthetic.
OS	Borrelia burgdorferi
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	473..1498
FT	/*tag= a
FT	/product= "P37/Flaa protein antigen"
FT	sig_peptide
FT	473..538
FT	/*tag= b
FT	mat_peptide
FT	539..1495
FT	/*tag= c
XX	
XX	WO9935272-A1.
PN	
XX	15-JUL-1999.
XX	
PD	
XX	06-JAN-1999;
XX	99WO-US00196.
XX	
XX	08-JAN-1998;
XX	98US-0004395.
XX	
PA	(INMR ) BIOMERIEUX INC.
XX	

Apple.

PI Gilmore RD, Johnson BJB;  
 XX WPI: 1999-430399/36.  
 DR P-PSDB; AAY28460.  
 XX  
 PT Recombinant FlaA/p37 protein antigen, useful in improved detection  
 of Borrelia burgdorferi  
 XX  
 PS Disclosure; Page 28; 34pp; English.  
 XX  
 CC The sequence encodes the recombinant protein antigen p37/FlaA. The  
 CC p37 protein product can be used in an assay for the detection of Borrelia  
 CC burgdorferi infection, the causative agent of Lyme disease. FlaA is  
 CC an outer sheath protein of the periplasmic flagella of B. burgdorferi.  
 CC FlaA is a prominent antigen in the early humoral immune response to  
 CC Borrelia burgdorferi infection, and significantly suitable for use in  
 CC improved serologic tests for exposure to Lyme disease spirochetes. The  
 CC detection of FlaA can augment the set of recombinant molecules that are  
 CC recognized early in the course of disease and contribute to improved  
 CC sensitivity of early testing for Lyme disease.  
 CC  
 XX Sequence 1655 BP; 560 A; 182 C; 308 G; 605 T; 0 other;

Query Match 98.4%; Score 1637; DB 20; Length 1655;  
 Best Local Similarity 99.5%; Pred. No. 3.2e-262;  
 Matches 1655; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 atgataatcttttcaaaaaggtttttatttcttctctagaagaggttgctga 60  
 Db 1 atgataatcttttcaaaaaggtttttatttcttctctagaagaggttgctga 60  
 QY 61 attaagatatgaagagagcaaaatttgaagaataattgaacatttgaataag 120  
 Db 61 attaagatatgaagagagcaaaatttgaagaataattgaacatttgaataag 120  
 QY 121 attctcttgaagaagaatgaggggttctcttgggttttatttaagtgatcaaat 180  
 Db 121 attctcttgaagaagaatgaggggttctcttgggttttatttaagtgatcaaat 180  
 QY 181 ttctatatcttttttaaaagaatgaggtttattatagtaaatcttcaagaagaatttatg 240  
 Db 181 ttctatatcttttttaaaagaatgaggtttattatagtaaatcttcaagaagaatttatg 240  
 QY 241 attcttttaataatgagatcaatcaatcttcttgatgtaaggtcaactcttttgcta 300  
 Db 241 attcttttaataatgagatcaatcaatcttcttgatgtaaggtcaactcttttgcta 300  
 QY 301 tgcctttaataaaacaatgaggttcaactatcctggttaagataaaagaattatata 360  
 Db 301 tgcctttaataaaacaatgaggttcaactatcctggttaagataaaagaattatata 360  
 QY 361 ttgtttgaaggtgtatcttaaaagagcaaggttgataaatttactataaataat 420  
 Db 361 ttgtttgaaggtgtatcttaaaagagcaaggttgataaatttactataaataat 420  
 QY 421 aattaaaaacggaataatttataaagatttataataaagaggtttgattacatgaaag 480  
 Db 421 aattaaaaacggaataatttataaagatttataataaagaggtttgattacatgaaag 480  
 QY 481 gaaagctaaagatatttattttttattatatacactgtcttcttgcctcaagaactga 540  
 Db 481 gaaagctaaagatatttattttttattatatacactgtcttcttgcctcaagaactga 540  
 QY 541 tggattagcagaggttctctaaagggcagagcctgagagaattagttcttaagatttgcga 600  
 Db 541 tggattagcagaggttctctaaagggcagagcctgagagaattagttcttaagatttgcga 600  
 QY 601 gcttgcaagagatccaagtccaactagacttgatcttaacaattatgttatatgata 660  
 Db 601 gcttgcaagagatccaagtccaactagacttgatcttaacaattatgttatatgata 660  
 QY 661 ttcgagcgtcttgatattgtaagccgagaagatagttgtagatcttgagataataa 720

Db 661 ttcgagcgtcttgatattgtaagccgagaagatagttgtagatcttgagataataa 720  
 QY 721 ttgagagcttttacttactctctctgaaggttgagaggttaagtaaaattcagttg 780  
 Db 721 ttgagagcttttacttactctctctgaaggttgagaggttaagtaaaattcagttg 780  
 QY 781 tgcgcgcgtcttgtaagagtagtcaaaaaggtacgcaggtgatactatttgggggt 840  
 Db 781 tgcgcgcgtcttgtaagagtagtcaaaaaggtacgcaggtgatactatttgggggt 840  
 QY 841 aagagtttgcttccaagctatctccaatcactgctgataattatgaccactttaaat 900  
 Db 841 aagagtttgcttccaagctatctccaatcactgctgataattatgaccactttaaat 900  
 QY 901 tccctttattcagggaaggtgcaatcaatttttaaggaaaggtcttattatatacat 960  
 Db 901 tccctttattcagggaaggtgcaatcaatttttaaggaaaggtcttattatatacat 960  
 QY 961 taaacccatgaagaataatgaagatctgttataagtttaggggtatgagataacttga 1020  
 Db 961 taaacccatgaagaataatgaagatctgttataagtttaggggtatgagataacttga 1020  
 QY 1021 ggtttatttgaagatagaaatgagcatggaatagcttactatggtgaactttaagtt 1080  
 Db 1021 ggtttatttgaagatagaaatgagcatggaatagcttactatggtgaactttaagtt 1080  
 QY 1081 taaaggttgagctgatttaatttggtaaatcccaatattcctcaatatacatccag 1140  
 Db 1081 taaaggttgagctgatttaatttggtaaatcccaatattcctcaatatacatccag 1140  
 QY 1141 aattatgaagaagatgcttccaatatacctctgcttcaagaataatgagatttaagc 1200  
 Db 1141 aattatgaagaagatgcttccaatatacctctgcttcaagaataatgagatttaagc 1200  
 QY 1201 tttaaggttccaaggtcacacagttccaagaatgtaaaatttcaatcttattatgaaga 1260  
 Db 1201 tttaaggttccaaggtcacacagttccaagaatgtaaaatttcaatcttattatgaaga 1260  
 QY 1261 tttaaggttcttattagatagcttaagttgttcaaatagattctgataattgacgttga 1320  
 Db 1261 tttaaggttcttattagatagcttaagttgttcaaatagattctgataattgacgttga 1320  
 QY 1321 tgcatttaaaaggtttagagctagcgaactgaatcccttcgttaaatgaaagcacaaga 1380  
 Db 1321 tgcatttaaaaggtttagagctagcgaactgaatcccttcgttaaatgaaagcacaaga 1380  
 QY 1381 aacttttaaaaggtttaaagcttagaagaataattctatcgcgtgaaggtcttcca 1440  
 Db 1381 aacttttaaaaggtttaaagcttagaagaataattctatcgcgtgaaggtcttcca 1440  
 QY 1441 aaacttggtaaaagattgagaggtgaaaaacctgaagaatcctccgaaaatttagt 1500  
 Db 1441 aaacttggtaaaagattgagaggtgaaaaacctgaagaatcctccgaaaatttagt 1500  
 QY 1501 ttaaatatcatgttaaaagctcaactaaaggtttgctt-----acataataaaataa 1552  
 Db 1501 ttaaatatcatgttaaaagctcaactaaaggtttgctt-----acataataaaataa 1552  
 QY 1561 taagaaatagatataagaaatattagatttgaagaatgaagaagctttagagattttt 1620  
 Db 1561 taagaaatagatataagaaatattagatttgaagaatgaagaagctttagagattttt 1620  
 QY 1621 tgaagaagctcaaaatctgttaatatccttgaagaagaatt 1663  
 Db 1621 tgaagaagctcaaaatctgttaatatccttgaagaagaatt 1663  
 QY 1663 tgaagaagctcaaaatctgttaatatccttgaagaagaatt 1663  
 Db 1663 tgaagaagctcaaaatctgttaatatccttgaagaagaatt 1663

RESULT 2  
 AAX61843  
 ID AAX61843 standard; DNA; 1011 BP.  
 XX  
 AC AAX61843;

Query Match	Similarity	98.5%	Score	98.6	DB	20	Length	1011
Best Local	Similarity	98.5%	Pred.	No.2	7e-15			
Matches	995	Conservative	0	Matches	15	Indels	0	Gaps
QY	489	aaagatattatattttttatattatccacgctctctttgtctcaagagacgtatggaattag	548					
Db	2	aaagatattatattttttatattatccacgctctctttgtctcaagagacgtatggaattag	61					
QY	549	cagaggttcttaaaagcgacagacgtctggaattagtcctagaattcttgcgcgagcttgcaa	608					
Db	62	cagaggttcttaaaagcgacagacgtctggaattagtcctagaattcttgcgcgagcttgcaa	121					
QY	609	ggagatccaaagttccaactgaacttgatctctcaaaattatgttgaattgtatattcggcg	668					
Db	122	ggagatccaaagttccaactgaacttgatctctcaaaattatgttgaattgtatattcggcg	181					
QY	669	ctctcgtgattgtttaagccggagagatagtggttgtagacttggagataaataattggaagc	728					
Db	182	ctctcgtgattgtttaagccggagagatagtggttgtagacttggagataaataattggaagc	241					
QY	729	tttctaactacctctctcgtcaaggtgtgcaggtctacgtttaaaaattcaagtgtgtgcgccg	788					
Db	242	tttctaactacctctctcgtcaaggtgtgcaggtctacgtttaaaaattcaagtgtgtgcgccg	301					
QY	789	ccgtgtgttaagagtgagttcaaaaaggtacgcaggtgatatactatttggggtaagattc	848					
Db	302	ccgtgtgttaagagtgagttcaaaaaggtacgcaggtgatatactatttggggtaagattc	361					

QY	849	tgtttccaagatattcccaatcattctcgctatgatattatgtccaccatttaaatctctttt	908
Db	362	tgtttccaagatattcccaatcattctcgctatgatattatgtccaccatttaaatctctttt	421
OY	909	attccgggaaagatgagcaatcaatttlttagcgaaaggtcttatgtataacttaaaacca	968
Db	422	attccgggaaagatgagcaatcaatttlttagcgaaaggtcttatgtataacttaaaacca	481
OY	969	tgaagaataatgaagtaactcgtttcatagtttagggatagagaatccttgaggctttat	1028
Db	482	tgaagaataatgaagtaactcgtttcatagtttagggatagagaatccttgaggctttat	541
OY	1029	ttgaagatataagatggaatggaatgatgcttatctctatggatattccttaagtttaagggt	1088
Db	542	ttgaagatataagatggaatggaatgatgcttatctctatggatattccttaagtttaagggt	601
OY	1089	ggcgctgatttaatttggccaactcctaactatatcttctaataatcacccagaattatla	1148
Db	602	ggcgctgatttaatttggccaactcctaactatatcttctaataatcacccagaattatla	661
OY	1149	aagagcatgttccaataattatccctctgcttcaagtaaaatgagatttaaggcttttagag	1208
Db	662	aagagcatgttccaataattatccctctgcttcaagtaaaatgagatttaaggcttttagag	721
OY	1209	tttcaaaagtcacagatccaagaagttaaaaattctaccttatgtttaaagatttaagag	1268
Db	722	tttcaaaagtcacagatccaagaagttaaaaattctaccttatgtttaaagatttaagag	781
OY	1269	tctcttatgataagctgaagtgtttccaatagatctgataatgacagtgagtcgtatla	1328
Db	782	tctcttatgataagctgaagtgtttccaatagatctgataatgacagtgagtcgtatla	841
OY	1329	aagtttatgagacagcaggaactgaaatcccttcgtataatttaaaggcacagaacttla	1388
Db	842	aagtttatgagacagcaggaactgaaatcccttcgtataatttaaaggcacagaacttla	901
OY	1389	aaagagttttaaagcttagagaaaaaattctatcgtctgaaagcctcttccaaaaacttgg	1448
Db	902	aaagagttttaaagcttagagaaaaaattctatcgtctgaaagcctcttccaaaaacttgg	961
OY	1449	tagaaaaagatgagagtgataaaacccgagaagaatcatcttcgaaaaatttag	1498
Db	962	tagaaaaagatgagagtgataaaacccgagaagaatcatcttcgaaaaatttag	1011
RESULT 3			
AAK61844			
ID	AAK61844 standard; DNA: 945 BP.		
XX	AAK61844;		
AC	19-JUL-1999 (first entry)		
XX	B. burgdorferi antigenic protein coding sequence, t07A.nt BB023.		
DT	Antigenic protein; vaccine; Lyme disease; Infection; detection; ss.		
XX	Borrelia burgdorferi.		
OS	MO9859071-A1.		
XX	18-JUN-1998; 98MO-US12718.		
XX	03-SEP-1997; 97US-0057483.		
PR	20-JUN-1997; 97US-0050359.		
PR	22-JUL-1997; 97US-0053344.		
PR	22-JUL-1997; 97US-0053377.		
XX	(HOMA-) HUMAN GENOME SCI INC.		
PA	(MEDI-) MEDIMMUNE INC.		
XX			



```
DB 20 atgataatcttttccaaaaggcttttatttccatcttcagcaaggatctgtgcta 79
QY 61 atttaagataatttaagagatgacaacaaatttgaagaataattagatcctttagtaaaag 120
DB 80 atttaagataatttaagagatgacaacaaatttgaagaataatttagatcctttagtaaaag 139
QY 121 atttcttcttaaggaagaatgaggggttctcctttgggttttaataagtgatccaagat 180
DB 140 atttcttcttaaggaagaagaagggttctcctttgggttttaataagtgatccaagat 199
QY 181 tttaatactctttttaaagaagaatgaggttattatgtaaatcttccaaggaatttatg 240
DB 200 tttaatactcttcttcaaaagaatgaggttataatgtaaatcttccaaggaatttatg 259
QY 241 attcctttaataatggtgata-----taataatcttttgatgcttaagtgcaatc 291
DB 260 attcctttaataatggtgataatgataatgataatgataatcttcttgatgcttaagtgcaatc 319
QY 292 ttttctgatagtctttaaataaaacaatgcgctttaactatcctgtgtaagaataaaaga 351
DB 320 ttttctgatagtctttaaataaaacaatgcgctttaactatcctgtgtaagaataaaaga 379
QY 352 ttattatctctgttgaagggtgatactttaaaggagcaagttga 395
DB 380 ttgtattctctgttgaagggtgatactttaaaggagcaagttga 423
```

## RESULT 5

AAK61739  
ID AAK61739 standard; DNA; 474 BP.

XX AAK61739;

XX 19-JUL-1999 (first entry)

DE B. burgdorferi antigenic protein coding sequence, f216.nt.

XX Antigenic protein; vaccine; Lyme disease; infection; detection; ss.

XX Borrelia burgdorferi.

XX MO9859071-A1.

PD 30-DEC-1998.

XX 18-JUN-1998; 98MO-US12718.

XX 03-SEP-1997; 97US-0057483.

PR 20-JUN-1997; 97US-0050359.

PR 22-JUL-1997; 97US-0053344.

PR 22-JUL-1997; 97US-0053377.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (MEDI-) MEDIMMUNE INC.

PI Choi GH, Erwin AL, Hanson MS, Iathigra R;

DR WPI: 1999-189980/16.

XX P-PsDB: AAY20042.

XX New isolated Borrelia burgdorferi nucleic acids - used to develop

XX products for the diagnosis, prevention and treatment of diseases

XX caused by Borrelia, particularly Lyme disease

XX Claim 1: Page 171; 275pp; English.

CC be used for detection of members of the Borrelia genus.

XX Sequence 474 BP; 145 A; 36 C; 85 G; 208 T; 0 other;

XX Query Match 22.4%; Score 372.8; DB 20; Length 474;

XX Best Local Similarity 97.3%; Pred. No. 2e-53; Indels 9; Gaps 1;

XX Matches 393; Conservative 0; Mismatches 2; Indels 9; Gaps 1;

```
QY 1 atgataatcttttccaaaaggcttttatttccatcttcagcaaggatctgtgcta 60
DB 71 atgataatcttttccaaaaggcttttatttccatcttcagcaaggatctgtgcta 130
QY 61 atttaagataatttaagagatgacaacaaatttgaagaataatttagatcctttagtaaaag 120
DB 131 atttaagataatttaagagatgacaacaaatttgaagaataatttagatcctttagtaaaag 190
QY 121 atttcttcttaaggaagaatgaggggttctcctttgggttttaataagtgatccaagat 180
DB 191 atttcttcttaaggaagaagaagggttctcctttgggttttaataagtgatccaagat 250
QY 181 tttaatactcttctttaaagaagaatgaggttataatgtaaatcttccaaggaatttatg 240
DB 251 tttaatactcttctttaaagaagaatgaggttataatgtaaatcttccaaggaatttatg 310
QY 241 attcctttaataatggtgata-----taataatcttttgatgcttaagtgcaatc 291
DB 311 attcctttaataatggtgataatgataatgataatgataatcttcttgatgcttaagtgcaatc 370
QY 292 ttttctgatagtctttaaataaaacaatgcgctttaactatcctgtgtaagaataaaaga 351
DB 371 ttttctgatagtctttaaataaaacaatgcgctttaactatcctgtgtaagaataaaaga 430
QY 352 ttattatctctgttgaagggtgatactttaaaggagcaagttga 395
DB 431 ttgtattctctgttgaagggtgatactttaaaggagcaagttga 474
```

## RESULT 6

AAF58252  
ID AAF58252 standard; DNA; 936 BP.

XX AAF58252;

XX 24-APR-2001 (first entry)

DE Oligonucleotide D1835.

XX Election-transfer group; ETM; mismatch; genotyping;

XX gene expression; ss.

XX Synthetic.

XX WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000MO-US20476.

XX 26-JUL-1999; 99US-0145695.

XX 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX Umek RM;

XX WPI: 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in

XX hybridization assays, e.g. for genotyping, allowing repeat analyses on

XX Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.

50 Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match	12.4%;	Score 207;	DB 22;	Length 936;
Best Local Similarity	1.1%;	Pred. No. 5e-26;		
Matches	9;	Conservative 555;	Mismatches 225;	Indels 0;
			Gaps	0;

[illegible]

AAAF58254  
ID AAAF58254 standard; DNA; 936 BP.

AC AAF58254

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1875.

KW Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.

OS Synthetic

PN WO200107665-A2

01-FEB-2001

PF 26-JUL-2000; 2000WO-US20476.

PR	26-JUL-1999;	99US-0145695.
PR	17-MAR-2000;	2000US-0190259

XX  
XX  
CLINICAL MICROSCOPES INC  
(CTIN-)

XX  
XX  
DT

XX WPB: 2001-159728/16  
 DP

Nucleic acids contain

PT hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface -

PS Example 6; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.

Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match	12.4%;	Score 207;	DB 22;	Length 936;
Best Local Similarity	1.1%;	Pred. No. 5e-26;		
Matches	9;	Conservative 555;	Mismatches 225;	Indels 0;
			Gaps	0;

[illegible]





Dt	24-APR-2001	(first entry)
xx	Oligonucleotide D2004.	
DE	Electron-transfer group; ETM; mismatch; genotyping;	
xx	gene expression; ss.	
KW	Synthetic.	
xx	WO200107665-A2.	
OS	01-FEB-2001.	
XX	26-JUL-2000; 2000WO-US20476.	
PX	26-JUL-1999; 99US-0145695.	
PR	17-MAR-2000; 2000US-0190259.	
xx	(CLIN-) CLINICAL MICRO SENSORS INC.	
PA	Umek RM;	
xx	WPI; 2001-159728/16.	
DR	Nucleic acids containing electron-transfer group, useful as labels in	
xx	hybridization assays, e.g. for genotyping, allowing repeat analyses on	
PT	a single surface -	
pt	-	
PS	Example 6; Page 128; 159pp; English.	
xx	The present invention relates to a composition comprising two nucleic	
CC	acids each containing an electron-transfer group (ETM) having	
CC	different redox potentials. The invention is used for electronic	
CC	detection of nucleic acids, especially of substitutions (mismatches)	
CC	and single-nucleotide polymorphisms, e.g. for genotyping,	
CC	monitoring gene expression.	
xx	Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;	
SO		

```

Query Match          12.4% Score 207; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 5e-26;
Matches    9; Conservative 555; Mismatches 225; Indels   0; Gaps   0.

Oy      1 atgataatcctttccaaaaagggttttatctcatcctagcaaggatttgtcta 60
        ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db       7 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 66

Oy      61 attaaagtattgaagaagtgaacaanaattgaaagataattagactttaagaag 120
        ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db       67 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 126

Oy      121 atttccttaagaaagaatgaggggttctcttggtttatlaagtgatccaagat 180
        ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db      127 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 186

Oy      181 ttcttatctttttaagaagtattatcgtaaacccttcaagagaatttag 240
        ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db      187 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 246

Oy      241 attccttaataatggtgatataatgaaccttttgatgtaagtcaacttttgcta 300
        ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db      247 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 306

Oy      301 tgccttaataaaaacaatgcgttaactaccgtaagtaagataaaaagattatc 360
        ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db      307 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 366

Oy      361 ttgttgaagggtgtacttaaaaggagcaagtgtgataaatltaactataaaaat 420
        ::::: ::::: ::::: ::::: || | ::::: ::::: ::::: :::::
Db      367 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 426

Oy      421 aattaaaaaacgaataattcataaagaattatatataagaaggttggttacatgaag 480

```

[illegible]



KM	gene expression; ss.
XX	
OS	Synthetic.
PN	WO200107665-A2.
XX	
PD	01-FEB-2001.
XX	
PF	26-JUL-2000; 2000WO-US20476.
XX	
PR	26-JUL-1999; 99US-0145695.
PR	17-MAR-2000; 2000US-0190259.
XX	
PA	(CLIN-) CLINICAL MICRO SENSORS INC.
XX	
PI	Umek RM;
XX	
DR	WPI: 2001-159728/16.
XX	
PT	Nucleic acids containing electron-transfer group, useful as labels in
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface
XX	
PS	Example 6; Page 127; 159pp; English.
XX	
CC	The present invention relates to a composition comprising two nucleic
CC	acids each containing an electron-transfer group (ETM) having
CC	different redox potentials. The invention is used for electronic
CC	detection of nucleic acids, especially of substitutions (mismatches)
CC	and single-nucleotide polymorphisms, e.g. for genotyping,
CC	monitoring gene expression.
XX	
QO	Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

```

Query Match Similarity 12.4% Score 207 DB 22 Length 938:
Best Local Similarity 1.1% Pred No. 55-26 Mismatches 225 Indels 0 Gaps 0
Matches 9 Conservative 555

QY 1 atgaataccttttcacaaaagtttttatttcattctctgcgaaggatttgtcta 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 66

QY 61 attaagatatttaagagatgaacaaatttggaaagataattgaatcttttagcaaaag 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 126

QY 121 attctctttaggaagcaatgagggcttctcttgggttttatttaagtgatccaagat 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 127 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 186

QY 181 ttatatattctttttaagaaatgagatttatatgaatacttcaagagaatttatg 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 246

QY 241 attctttaaataatgggatataatgaatctcttgaatgaagtccaacttttgcta 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 306

QY 301 tgcctttaaataaaacaaatgcgctttaactatccgcgtgaagataaaagaattatattc 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 366

QY 361 ttgttgaaaggtgatctcttaaagagacaaagtgtgataatcaattactactataaaaat 420
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 367 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 426

QY 421 aattcaaaaacgaaaattttaaaagatttatataaagagtggtttacatgaaag 480
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 427 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 486

QY 481 gaagctaaagtatatttatttttattatccacgcttcttcttgctcaagagactga 540
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 487 wwwwwwwww 546  
QY 541 tggattacagaggtcttaaaagcgagcctggagaattagcttagatttgcga 600  
Db 547 wwwwwwwww 606  
QY 601 gcttgcagagatccaagttcaactagacttgatcttaacaattagtgtatgata 660  
Db 607 wwwwwwwww 666  
QY 661 ttcggcgcttcgtgtatgttaagccggaagatatgttgcatacttggaataaa 720  
Db 667 wwwwwwwww 726  
QY 721 ttggagcgtttactactcctctgcagaggttcgagccttaacaaattcagttgt 780  
Db 727 wwwwwwwww 786  
QY 781 tgcgcgcgc 789  
Db 787 wgcgcgcgc 795

## RESULT 12

AAF58252/C

ID AAF58252 standard; DNA; 936 BP.

AC AAF58252;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1835.

KW Electron-transfer group; ETM; mismatch; genotyping;

KM gene expression; ss.

OS Synthetic.

PN WO200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

DR WPI: 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface

PS Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.

SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 12.4%; Score 206; DB 22; Length 936;  
Best Local Similarity 0.6%; Pred. No. 7.4e-26;  
Matches 5; Conservative 554; Mismatches 219; Indels 0; Gaps 0;

QY 1 atgatactttttcaaaaaggttttatttcaattcagcaaggagttgtgcta 60  
Db 778 wwwwwwwww 719  
QY 61 attaagattttaagagatgaacaaaatttgaaagataattgactcttagtaaaag 120  
Db 718 wwwwwwwww 659  
QY 121 attcttttaggaagcaatagagggtttcttttggtttcttaagtgatccaagt 180  
Db 658 wwwwwwwww 599  
QY 181 tttaattctttttaagaatgagttatlbtaaatcttcaagagaatttatg 240  
Db 598 wwwwwwwww 539  
QY 241 attctttaataatggtatcataatgaatcttttgatgtaagtcacatcttttgcta 300  
Db 538 wwwwwwwww 479  
QY 301 tgccttaataaaacaatgcgtttaactacactcgtgaagataaaaagattatc 360  
Db 478 wwwwwwwww 419  
QY 361 ttgttgaagggtgtatcttaaaaggacaaagttgataaattaacttaataaaat 420  
Db 418 wwwwwwwww 359  
QY 421 aattaaaaacgaaatttataaagaattatataagaagtggttcacagaaag 480  
Db 358 wwwwwwwww 299  
QY 481 gaaagctaaagatttatttttlatlaccactgtctcttttgcacagactga 540  
Db 298 wwwwwwwww 239  
QY 541 tggattacagaggttcttaaaaggcagagcctggagaattagcttagatttgcga 600  
Db 238 wwwwwwwww 179  
QY 601 gcttgcagagatccaagttcaactagacttgatcttaacaattagtgtatgata 660  
Db 178 wwwwwwwww 119  
QY 661 ttcggcgcttcgtgtatgttaagccggaagatatgttgcatacttggaataaa 720  
Db 118 wwwwwwwww 59  
QY 721 ttggagcgtttactactcctctgcagaggttcgagccttaacaaattcagtt 778  
Db 58 wwwwwwwww 1

## RESULT 13

AAF58254/C

ID AAF58254 standard; DNA; 936 BP.

AC AAF58254;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1875.

KW Electron-transfer group; ETM; mismatch; genotyping;

KM gene expression; ss.

OS Synthetic.

PN WO200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

XX 26-JUL-1999: 99US-0145695.  
PR 17-MAR-2000: 2000US-0190259.  
XX (CLIN-) CLINICAL MICRO SENSORS INC.  
XX Umek RM:  
XX WPI: 2001-159728/16.  
XX  
XX Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface  
XX  
XX Example 6: Page 127; 159pp; English.  
XX  
XX The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX  
XX  
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 12.4%; Score 206; DB 22; Length 936;  
Best Local Similarity 0.6%; Pred. No. 7.4e-26;  
Matches 5; Conservative 554; Mismatches 219; Indels 0; Gaps 0;

Oy 1 atgataatctttccaaaagggttttatttcatcttcagcaaggagttgtgcta 60  
Db 778 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 719  
Oy 61 attaagataatgaagatgacaataattgaagaataattagatctttgtaaaag 120  
Db 718 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 659  
Oy 121 atttcttttagaagcaatgaggtttcttcttggttttataagatcaaat 180  
Db 658 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 599  
Oy 181 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 240  
Db 598 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 539  
Oy 241 attcttataatgagtgatataatgattcttgatgtaagtcatttgcta 300  
Db 538 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 479  
Oy 301 tgccttataaataaagcgcttcaactcttcgtaagataaagataatctatc 360  
Db 478 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 419  
Oy 361 tgcgtgaagggtgatacttaaaaggagcaagtgataaatttactataaata 420  
Db 418 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 359  
Oy 421 aattaaaaacgaataattataaagattalataaagagtggtttacataagaag 480  
Db 358 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 299  
Oy 481 gaaagctaaagattatttttttattatccactgtctcttttgctcaagagctga 540  
Db 298 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 239  
Oy 541 tgcgtgaaggaggttcttaaaaggcgagcctgagaaattagcttagatttcgca 600  
Db 238 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 179  
Oy 601 gcttgcaagagatcgaagtccaactgactgacttacaataatgltgataatgata 660  
Db 178 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 119

Oy 661 ttccggcgcttctgtagttgtaagcgaagataatgtagttgtagatcttggaataaa 720  
Db 118 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 59  
Oy 721 ttggaagcgtttactactctcttcgcaaggttcgaaggttcaataaattcaat 778  
Db 58 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 1

RESULT 14  
ID AAF58257 standard; DNA; 936 BP.  
XX AAF58257;  
XX

XX 24-APR-2001 (first entry)  
XX  
XX Oligonucleotide D1954.  
XX

XX Electron-transfer group; ETM; mismatch; genotyping;  
KW gene expression; ss.  
XX  
XX Synthetic.  
XX

XX MO200107665-A2.  
XX

XX 01-FEB-2001.  
XX

XX 26-JUL-2000; 2000WO-US20476.  
XX

XX 26-JUL-1999: 99US-0145695.  
PR 17-MAR-2000: 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.  
XX

XX Umek RM:  
XX

XX WPI: 2001-159728/16.  
XX

XX Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface  
XX

XX Example 6: Page 127; 159pp; English.  
XX

XX The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.  
XX

SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 12.4%; Score 206; DB 22; Length 936;  
Best Local Similarity 0.6%; Pred. No. 7.4e-26;  
Matches 5; Conservative 554; Mismatches 219; Indels 0; Gaps 0;

Oy 1 atgataatcttttcaaaaagggttttatttcatcttcagcaaggagttgtgcta 60  
Db 778 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 719  
Oy 61 attaagataatgaagatgacaataattgaagaataattagatctttgtaaaag 120  
Db 718 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 659  
Oy 121 atttcttttagaagcaatgaggtttcttcttggttttataagatcaaat 180  
Db 658 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 599  
Oy 181 ttttataatcttttcaaaagggttttatttcatcttcagcaaggagttgtgcta 240

```
DB 598 WWWW... 539
QY 241 attccttaataatgataatgaatccttgatgtaagtaacatttgcta 300
DB 598 WWWW... 479
QY 301 tgccttaataaacaatgcgttaactatccgtgaagtaaaagatatatc 360
DB 478 WWWW... 419
QY 361 ttgtgaaggtgtatcttaaaagagcaagtgtgaaataatcttaataaaat 420
DB 418 WWWW... 359
QY 421 aattaaaaacgaatattataaagattatataaaggttggttacaagaag 480
DB 358 WWWW... 299
QY 481 gaaagctaaagtattatttttttattatcacactgtcttctgccaagagactga 540
DB 298 WWWW... 239
QY 541 tggattagcagaagggtcttaaaagagcagccttgagagaattagcttagatttgcga 600
DB 238 WWWW... 179
QY 601 gcttgaagagatccaagtccaactagacttgatcttaacaattatgttattgata 660
DB 178 WWWW... 119
QY 661 ttcgagcgcttcgttatgttaagccggaagatagtgtgtagatcttggaataaa 720
DB 118 WWWW... 59
QY 721 ttgagcgcttactactcctctgcgaaggttcgaaggttactgtaaaattcagtt 778
DB 58 WWWW... 1

RESULT 15
AAFS8259/C
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PE 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI umek RM;
XX
DR WPI: 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
```

```
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 12.4%; Score 206; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 7 4e-26;
Matches 5; Conservative 554; Mismatches 219; Indels 0; Gaps 0;

QY 1 atgataatcttttcaaaaaggttttatttcttctagcaaggattgtgcta 60
DB 778 WWWW... 719
QY 61 attaagatattgaagatgaacaaattgaaagataattagatcttttagtaaaag 120
DB 718 WWWW... 659
QY 121 attccttttaggaagcaatgaggggttcttcttgggttttattatgaatgaat 180
DB 658 WWWW... 599
QY 181 tttaattctttttaagaatgagatttattatgaatcttcaagagaatttag 240
DB 598 WWWW... 539
QY 241 attcctttaataatggtatataatgaatccttgatgtaaggtcaacttttgcta 300
DB 538 WWWW... 479
QY 301 tgccttaataaacaatgcgttaactatccgtgaagtaaaagatatatc 360
DB 478 WWWW... 419
QY 361 ttgtgaaggtgtatcttaaaagagcagccttgagagaattagcttagatttgcga 600
DB 418 WWWW... 359
QY 601 gcttgaagagatccaagtccaactagacttgatcttaacaattatgttattgata 660
DB 178 WWWW... 119
QY 661 ttcgagcgcttcgttatgttaagccggaagatagtgtgtagatcttggaataaa 720
DB 118 WWWW... 59
QY 721 ttgagcgcttactactcctctgcgaaggttcgaaggttactgtaaaattcagtt 778
DB 58 WWWW... 1

Search completed: March 20, 2002, 01:57:36
Job time: 7017 sec
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2002, 01:57:36 ; Search time 1691.33 Seconds  
(Without alignments)  
10.645 Million cell updates/sec

Title: US-09-004-395-3  
Perfect score: 21  
Sequence: 1 atgaaaggaagctaaagt 21

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq.1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	20	AXX89521
2	21	100.0	21	20	AXX89520
3	17.4	82.9	626	21	AAC49540
4	17.4	82.9	627	21	AAC50183
5	17.4	82.9	678	21	AAC34709
6	16.8	80.0	383	22	AA135657
7	16.8	80.0	383	22	AA104115
8	16.8	80.0	407	22	AA166213
9	16.8	80.0	476	22	AA111559
10	16.8	80.0	476	22	AA132844
11	16.8	80.0	476	22	AA101480

12	16.8	80.0	2569	20	AXX40071	Colon cancer assoc
13	16.4	78.1	222	21	AA229782	Purative PMG-1 fro
14	16.4	78.1	346	22	AAH83805	Human ovarian tumo
15	16.4	78.1	346	22	AAH83806	Human ovarian tumo
16	16.4	78.1	512	22	AAH12296	Human cDNA clone (
17	16.4	78.1	525	22	AAH83296	Human ovarian tumo
18	16.4	78.1	683	22	AA117578	Human breast cance
19	16.4	78.1	687	22	AA144891	Human breast cance
20	16.4	78.1	760	21	AA142519	Arabidopsis thalia
21	16.4	78.1	769	20	AA197936	Human secreted pro
22	16.4	78.1	803	21	AA151986	Arabidopsis thalia
23	16.4	78.1	1257	21	AAA75110	cDNA encoding a hu
24	16.4	78.1	1257	21	AAA37144	Human PRO1550 cDNA
25	16.4	78.1	1257	22	AA192118	Human PRO1550 cDNA
26	16.4	78.1	1257	22	AA154507	Probe #60 used in
27	16.4	78.1	1286	21	AA126456	Human secreted pro
28	16.4	78.1	1288	20	AA184525	Human secreted pro
29	16.4	78.1	1295	22	AA199687	Human protein enco
30	16.4	78.1	1342	21	AA126424	Human secreted pro
31	16.4	78.1	1619	21	AA129780	Sequence upstream
32	16.4	78.1	2002	22	AA114767	Human cDNA sequenc
33	16.4	78.1	130480	22	AA125833	R. marinus bacteri
34	16.2	77.1	26	21	AA170952	Single nucleotide
35	16.2	77.1	26	21	AA170955	Single nucleotide
36	16.2	77.1	26	21	AA170958	Single nucleotide
37	16.2	77.1	96	21	AA123158	Human secreted pro
38	16.2	77.1	238	21	AA170953	Single nucleotide
39	16.2	77.1	238	21	AA170956	Single nucleotide
40	16.2	77.1	238	21	AA170959	Single nucleotide
41	16.2	77.1	332	22	AA172390	Human cervical can
42	16.2	77.1	337	22	AA170840	Human cervical can
43	16.2	77.1	365	22	AA169505	Human cervical can
44	16.2	77.1	480	20	AA113749	Enterococcus faeca
45	16.2	77.1	575	22	AA172968	Human cervical can

#### ALIGNMENTS

RESULT 1  
ID AAX89521 standard; DNA: 21 BP.  
AC AAX89521:  
XX 19-OCT-1999 (first entry)  
XX PCR primer F1 for amplification of FlaA.  
XX P37; FlaA; Lyme disease; antigen; assay; PCR primer; F1 construct; ss.  
XX Synthetic.  
XX Borrelia burgdorferi.  
XX W09935272-A1.  
XX 15-JUL-1999.  
XX 06-JAN-1999; 99WO-US00196.  
XX 08-JAN-1998; 98US-0004395.  
XX (INMR) BIOMERIEUX INC.  
XX Gilmore RD, Johnson BJB;  
XX WPI; 1999-43039/36.  
XX Recombinant FlaA/P37 protein antigen, useful in improved detection  
XX of Borrelia burgdorferi  
XX Example 2; Page 11; 34pp; English.

CC The sequence is a PCR primer denoted F1, used to create the F1  
CC construct, which consists of the entire coding sequence of P37  
CC (AA89520). The P37 protein product can be used as an assay for the  
CC detection of Borrelia burgdorferi infection, the causative agent of Lyme  
CC disease. Flaa is an outer sheath protein of the periplasmic flagella of  
CC B.burgdorferi. Flaa is a prominent antigen in the early humoral immune  
CC response to Borrelia burgdorferi infection, and is significantly suitable  
CC for use in improved serologic tests for exposure to Lyme disease  
CC Spirochetes. The detection of Flaa can augment the set of recombinant  
CC molecules that are recognized early in the course of disease and  
CC contribute to improved sensitivity of early testing for Lyme disease.  
CC  
SQ Sequence 21 BP; 12 A; 1 C; 5 G; 3 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 21;  
Best Local Similarity 100.0%; Pred. No. 4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atgaagaagaaagctaaagt 21  
|||||  
Db 1 atgaagaagaaagctaaagt 21

RESULT 2  
AA89520  
ID AA89520 standard; DNA; 1655 BP.

XX  
AC AA89520;

DT 19-OCT-1999 (first entry)

XX Nucleic acid sequence encoding the Flaa gene product.

KW P37; Flaa; Lyme disease; antigen; assay; spirochete; ss.

OS Synthetic.

OS Borrelia burgdorferi

FT Key Location/Qualifiers  
FT CDS 473..1498  
FT /\*tag= a

FT sig\_peptide /product= "P37/Flaa protein antigen"

FT /\*tag= b 473..538

FT mat\_peptide 539..1495

FT /\*tag= c

PN WO935272-A1.

PD 15-JUL-1999.

PF 06-JAN-1999; 99WO-US00196.

PR 08-JAN-1998; 98US-0004395.

PA (INNR) BIOMERIEUX INC.

PI Gilmore RD, Johnson BJB.

DR WPI; 1999-430399/36.

DR P-PSDB; AAV28460.

PT Recombinant Flaa/P37 protein antigen, useful in improved detection

PT of Borrelia burgdorferi

PS Disclosure; Page 28; 34pp; English.

CC The sequence encodes the recombinant protein antigen P37/Flaa. The  
CC P37 protein product can be used in an assay for the detection of Borrelia  
CC burgdorferi infection, the causative agent of Lyme disease. Flaa is  
CC an outer sheath protein of the periplasmic flagella of B. burgdorferi.  
CC Flaa is a prominent antigen in the early humoral immune response to

CC Borrelia burgdorferi infection, and significantly suitable for use in  
CC improved serologic tests for exposure to Lyme disease spirochetes. The  
CC detection of Flaa can augment the set of recombinant molecules that are  
CC recognized early in the course of disease and contribute to improved  
CC sensitivity of early testing for Lyme disease.  
CC  
SQ Sequence 1655 BP; 560 A; 182 C; 308 G; 605 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 1655;  
Best Local Similarity 100.0%; Pred. No. 4.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atgaagaagaaagctaaagt 21  
|||||  
Db 473 atgaagaagaaagctaaagt 493

RESULT 3  
AAC49540/C  
ID AAC49540 standard; DNA; 626 BP.

XX  
AC AAC49540;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 61536.

DE Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PM EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 29-MAR-1999; 99US-0126264.

PR 25-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132486.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 19-MAY-1999; 99US-0134768.

PR 20-MAY-1999; 99US-0134941.

PR 21-MAY-1999; 99US-0135124.

PR 24-MAY-1999; 99US-0135353.

PR 25-MAY-1999; 99US-0135629.

PR 99US-0136021.



PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0138119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.

PR 29-OCT-1999; 99US-0162142.  
Query Match 82.9%; Score 17.4; DB 21; Length 626;  
Best Local Similarity 94.7%; Pred. No. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 gaaaggaagctaaagt 21  
|||||||  
Db 505 GAAAGGAAATGCTAAAGT 487  
RESULT 4  
AAC50183/c  
ID AAC50183 standard; DNA; 627 BP.  
XX  
AC AAC50183;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 63886.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
XX  
OS EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135628.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139859.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142380.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0143542.  
PR 13-JUL-1999; 99US-0143624.  
PR 14-JUL-1999; 99US-0144005.  
PR 15-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147028.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0158765.  
PR 07-OCT-1999; 99US-0158829.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 82.9%; Score 17.4; DB 21; Length 627;  
Best Local Similarity 94.7%; Pred. No. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 3 gaaaggaagctaaagt 21

DB 507 GAAAGCATGCTAAAGT 489  
|||||||  
RESULT 5  
AAC34709/c  
ID AAC34709 standard; DNA; 678 BP.  
XX  
AC AAC34709;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 7604.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123348.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.

```
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147121.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158329.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      82.9%; Score 17.4; DB 21; Length 678;
Best Local Similarity 94.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
RESULT 6
AAI35657
ID AAI35657 standard; DNA; 383 BP.
```

```
XX AC AA135657;
XX DT 17-OCT-2001 (first entry)
XX DE Probe #4343 used to measure gene expression in human placenta sample.
XX KW Probe: microarray; human; placenta; antenatal diagnosis;
XX KM genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488897/53.
XX DT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 25; SEQ ID NO 4343; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 383 BP; 127 A; 69 C; 69 G; 118 T; 0 other;

Query Match      80.0%; Score 16.8; DB 22; Length 383;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgaagaagaaagctaaagt 21
   ||||||| | |||||||
DB 145 tgaagaagtagctaaagt 164

RESULT 7
AA104115
ID AA104115 standard; DNA; 383 BP.
XX AC AA104115;
XX DT 09-OCT-2001 (first entry)
XX DE Probe #4106 used to measure gene expression in human breast sample.
XX KW Probe: human; breast disease; breast cancer; development disorder; ss;
XX KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN WO200157270-A2.
XX PD 09-AUG-2001.
```

```
XX XX 29-JAN-2001; 2001WO-US00661.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-476286/51.
XX DT Novel single exon nucleic acid probe used to measuring gene expression
XX PT in a human breast -
XX PS Claim 25; SEQ ID NO 4106; 322pp; English.
XX CC The present invention relates to novel single exon nucleic acid probes.
XX CC The present sequence is one such probe. The probes are useful for
XX CC measuring human gene expression in a human breast sample, where the probe
XX CC hybridises at high stringency to a nucleic acid expressed in the human
XX CC breast. The probes are useful for predicting, diagnosing, grading,
XX CC staging, monitoring and prognosing diseases of the human breast,
XX CC particularly those diseases with polygenic aetiology. The diseases
XX CC include: breast cancer, disorders of development, inflammatory diseases
XX CC of the breast, fibrocystic changes, proliferative breast disease and
XX CC non-carcinoma tumours.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pcl_sequences.
XX SQ Sequence 383 BP; 127 A; 69 C; 69 G; 118 T; 0 other;

Query Match      80.0%; Score 16.8; DB 22; Length 383;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgaagaagaaagctaaagt 21
   ||||||| | |||||||
DB 145 tgaagaagtagctaaagt 164

RESULT 8
AA166213/C
ID AA166213 standard; cDNA; 407 BP.
XX AC AA166213;
XX DT 09-APR-2001 (first entry)
XX DE Novel human polynucleotide, SEQ ID NO: 1969.
XX KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
XX KM breast cancer; lung cancer; cancer detection; ss.
XX OS Homo sapiens.
XX PN WO200102568-A2.
XX PD 11-JAN-2001.
XX PF 30-JUN-2000; 2000WO-US18374.
XX PR 02-JUL-1999; 99US-0142310.
XX PR 02-JUL-1999; 99US-0142311.
XX PA (CHIR ) CHIRON CORP.
```

PA (HYSEQ-)HYSEQ INC.  
 XX  
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A,  
 PI Reinhard C, Randazzo P, Kennedy GC, Pot D, Lamson G, Dirmanc R;  
 PI Cirenjakov R, Dirmanc S, Dickson M, Labat I, Leshkowitz D;  
 PI Kita D, Garcia V, Jones LW, Strache-Grain B;  
 XX  
 DR WP1; :2001-091805/10.  
 XX  
 PT Library of polynucleotides for diagnosing a cancerous state of a  
 PT mammalian cell and detecting cancer, particularly of the colon or  
 PT prostate, comprises 3351 human polynucleotide sequences -  
 XX  
 XX  
 Claim 9; Page 828; 1046pp; English.  
 XX  
 XX The present sequence is one of 3351 sequences in a library of human  
 CC polynucleotides. The library is used to detect differentially expressed  
 CC genes correlated with a cancerous state of a mammalian cell and can  
 CC detect colon, prostate, breast and lung cancer. The library can be used  
 CC to produce probes for detection of mRNA and to produce additional copies  
 CC of the polynucleotides. The probes can be used for chromosome mapping of  
 CC the polynucleotide and for detection of transcription levels. Ribozymes  
 CC or antisense oligonucleotides can be generated. The polynucleotides and  
 CC their gene products are used as genetic or biochemical markers (e.g. in  
 CC blood or tissues) that will detect the earliest changes along the  
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and  
 CC preventative interventions. The polynucleotides, polypeptides and  
 CC antibodies against them can be used in pharmaceutical compositions to  
 CC treat the cancers and proliferative disorders such as neoplasia,  
 CC dysplasia and hyperplasia.  
 XX  
 XX Sequence 407 BP; 109 A; 62 C; 87 G; 149 T; 0 other;

```

Query Match      80.0%; Score 16.8; DB 22; Length 407;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0

OY      2  Tgaaaagaaagctaaagt 21
        |||||  |||||
DB      52  TCGAAATATAAACTTAAAGT 33

RESULT  9
ID      AAt11559
XX      AAt11559 standard; DNA: 476 BP.
AC      AAt11559;
XX      12-OCT-2001 (first entry)
DT      XX
DE      Probe #1492 for gene expression analysis in human cervical cell sample.
XX      DE
XX      Probe; human; microarray; gene expression; cervical epithelial cell;
KW      cervical cancer; ss.
XX      KW
XX      Homo sapiens.
OS      Homo sapiens.
XX      OS
XX      W0200157278-A2.
PN      W0200157278-A2.
XX      PN
XX      09-AUG-2001.
PD      09-AUG-2001.
XX      PD
XX      30-JAN-2001; 2001WO-US00670.
PF      30-JAN-2001; 2001WO-US00670.
XX      PF
XX      04-FEB-2000; 2000US-0180312.
PR      04-FEB-2000; 2000US-0207456.
XX      PR
XX      26-MAY-2000; 2000US-0608408.
PR      26-MAY-2000; 2000US-0608408.
XX      PR
XX      30-JUN-2000; 2000US-0632366.
PR      30-JUN-2000; 2000US-0632366.
XX      PR
XX      03-AUG-2000; 2000US-0234687.
PR      03-AUG-2000; 2000US-0234687.
XX      PR
XX      21-SEP-2000; 2000US-0236359.
PR      21-SEP-2000; 2000US-0236359.
XX      PR
XX      27-SEP-2000; 2000US-0024263.
PR      27-SEP-2000; 2000US-0024263.
XX      PR
XX      04-OCT-2000; 2000US-0024263.
PR      04-OCT-2000; 2000US-0024263.
XX      PR
XX      (MOLE-) MOLECULAR DYNAMICS INC.
PR      (MOLE-) MOLECULAR DYNAMICS INC.

```

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-488901/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human cervical epithelial cells -  
PT  
XX  
PS Claim 25; SEQ ID NO 1492; 487pp; English.  
XX  
XX The present invention relates to human single exon nucleic acid probes  
CC (SENP). The present sequence is one such probe. The SNPs are derived  
CC from human HeLa cells. The SENPs can be used to produce a single exon  
CC microarray, which can be used for measuring human gene expression in a  
CC sample derived from human cervical epithelial cells. By measuring gene  
CC expression, the probes are therefore useful in grading and/or staging  
CC of diseases of the cervix, notably cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 476 BP; 154 A; 79 C; 81 G; 162 T; 0 other;

Query Match	80.0%;	Score 16.8;	DB 22;	length 476;
Best Local Similarity	90.0%;	Pred. No. 2.4e+02;		
Matches	18;	Conservative	0;	Mismatches 2; Indels 0; Gaps 0;
Oy	2	tgaaaaggaagctaaagt	21	
Db	159	tgaaaagtgtagctaaagt	178	
RESULT	10			
AAI32844				
ID	AAI32844	standard; DNA; 476 BP.		
XX	AC	AAI32844;		
XX	DT	17-OCT-2001 (first entry)		
XX	DE	Probe #1530 used to measure gene expression in human placenta sample.		
XX	KW	Probe; microarray; human; placenta; antenatal diagnosis;		
XX	KW	genetic disorder; ss.		
XX	OS	Homo sapiens.		
XX	PN	WO200157272-A2.		
XX	PD	09-AUG-2001.		
XX	PF	30-JAN-2001; 2001WO-US00663.		
XX	PR	04-FEB-2000; 2000US-0180312.		
XX	PR	26-MAY-2000; 2000US-0207456.		
XX	PR	30-JUN-2000; 2000US-0608408.		
XX	PR	03-AUG-2000; 2000US-0633366.		
XX	PR	21-SEP-2000; 2000US-0234687.		
XX	PR	27-SEP-2000; 2000US-0236359.		
XX	PR	04-OCT-2000; 2000GB-0024263.		
XX	PA	(MOLE-) MOLECULAR DYNAMICS INC.		
XX	PI	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX	DR	WPI; 2001-488897/53.		
XX	PT	Human genome-derived single exon nucleic acid probes useful for		
XX	PS	analyzing gene expression in human placenta -		
XX	PS	Claim 25; SEQ ID NO 1530; 654bp; English.		

CC The present invention relates to single exon nucleic acid probes (SENP).  
CC The present sequence is one such probe. The probes are useful for  
CC producing a microarray for predicting, measuring and displaying gene  
CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders.

XX  
SQ Sequence 476 BP; 154 A; 79 C; 81 G; 162 T; 0 other;

Query Match 80.0%; Score 16.8; DB 22; Length 476;  
Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 tgaagaaggaagctaaagt 21  
||||||| |  
DB 159 tgaagaagctagctaaagt 178

RESULT 11

ID AAI01480 standard; DNA; 476 BP.

XX AAI01480;

DT 09-OCT-2001 (first entry)

DE Probe #1471 used to measure gene expression in human breast sample.

XX Probe; human; breast disease; breast cancer; development disorder; ss;

KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

PN WO200157270-A2.

PD 09-AUG-2001.

PF 29-JAN-2001; 2001WO-US00661.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632365.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression

XX in a human breast -

XX Claim 25; SEQ ID No 1471; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes.

XX The present sequence is one such probe. The probes are useful for

XX measuring human gene expression in a human breast sample, where the probe

XX hybridises at high stringency to a nucleic acid expressed in the human

XX breast. The probes are useful for predicting, diagnosing, grading,

XX staging, monitoring and prognosing diseases of the human breast,

XX particularly those diseases with polygenic aetiology. The diseases

XX include: breast cancer, disorders of development, inflammatory diseases

XX of the breast, fibrocystic changes, proliferative breast disease and

XX non-carcinoma tumours.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 476 BP; 154 A; 79 C; 81 G; 162 T; 0 other;

Query Match 80.0%; Score 16.8; DB 22; Length 476;  
Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 tgaagaaggaagctaaagt 21  
||||||| |  
DB 159 tgaagaagctagctaaagt 178

RESULT 12

ID AAX40071 standard; DNA; 2569 BP.

XX AAX40071;

DT 02-JUL-1999 (first entry)

DE Colon cancer associated gene.

XX Cancer associated antigen; diagnosis; research; treatment; human;

KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

XX prostate cancer; ss.

XX Homo sapiens.

PN WO904265-A2.

PD 28-JAN-1999.

PF 15-JUL-1998; 98WO-US14679.

PR 22-JUN-1998; 98US-0102322.

PR 17-JUL-1997; 97US-0896164.

PR 10-OCT-1997; 97US-0061599.

PR 10-OCT-1997; 97US-0061765.

PR 10-OCT-1997; 97US-0948705.

PR 11-OCT-1997; 97GB-0021697.

XX (LUDWIG) LUDWIG INST CANCER RES.

XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

XX Pfreundschuh M, Sahlin U, Scanlan MJ, Stockert E;

XX Tureci O;

XX WPI; 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides -

XX isolated using sera from cancer patients, used to develop products

XX for the diagnosis, monitoring or treatment of cancers

XX Claim 67; Page 677-678; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised

XX by expression of a human cancer associated antigen precursor coded for by

XX a nucleic acid molecule (NAM). The method comprises: (a) contacting a

XX biological sample isolated from a subject with an agent that specifically

XX binds to the NAM, an expression product or a fragment of an expression

XX product complexed with an HLA molecule; and (b) determining the

XX interaction between the agent and the NAM or the expression product as a

XX determination of the disorder. The products and methods can be used in

XX the diagnosis, monitoring, research, or treatment of conditions

XX characterised by the expression of various cancer associated antigens.

XX The invention provides nucleic acid sequences and encoded polypeptides

XX which are cancer associated antigen precursors expressed in human breast

XX cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

XX lung cancer.

XX Sequence 2569 BP; 865 A; 418 C; 564 G; 712 T; 10 other;

XX Query Match 80.0%; Score 16.8; DB 20; Length 2569;

Best Local Similarity 90.0%; Pred. No. 2.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 2 tgaagaagaagctaaagt 21  
Db 1242 tgaagaagaagctaaagt 1261

RESULT 13  
AA229782/C  
ID AA229782 standard; DNA: 222 BP.  
XX  
AC AA229782;  
XX  
DT 27-MAR-2000 (first entry)  
XX  
DE Putative PMG-1 from tomato mutant UQ14.  
XX  
KW Tomato mutant UQ14; sucrose synthase gene;  
KM PMGS; phenotype modifying genetic sequence; plant pathogen resistance;  
KM senescence timing; starch metabolism; ds.  
XX  
OS Lycopersicon esculentum.  
XX  
PN WO9963068-A1.  
XX  
PD 09-DEC-1999.  
XX  
PE 04-JUN-1999; 99WO-AU00434.  
XX  
PR 04-JUN-1998; 98AU-0003901.  
PR 04-JUN-1998; 98AU-0003903.  
PR 25-SEP-1998; 98AU-0006169.  
PR 25-SEP-1998; 98AU-0006174.  
XX  
PA (UYOU ) UNIV QUEENSLAND.  
XX  
PI Carrot 1 BJ;  
XX  
DR WPI: 2000-116368/10.  
XX  
PT New polynucleotides that increase gene expression in plants used to  
PT produce transgenic plants with resistance to plant pathogens -  
XX  
PS Claim 20; Page 63; 93pp; English.  
XX  
CC The present sequence is a phenotype modifying genetic sequence upstream  
CC of ds insertion. This shows homology to 3' untranslated region of potato  
CC sucrose synthase gene. This can be used to produce transgenic plants  
CC having altered phenotypic traits, such as resistance to plant pathogens,  
CC senescence timing, starch metabolism, cell growth, expansion and/or  
CC division, and the shape of cells, tissues or organs.  
XX  
SQ Sequence 222 BP; 63 A; 28 C; 41 G; 66 T; 24 other;

Query Match 78.1%; Score 16.4; DB 21; Length 222;  
Best Local Similarity 94.4%; Pred. No. 3.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 aaaaggaagctaaagt 21  
Db 152 AAAAGAAAAGCTAAAGT 135

RESULT 14  
AAH83805  
ID AAH83805 standard; cDNA: 346 BP.  
XX  
AC AAH83805;  
XX  
DT 25-SEP-2001 (first entry)  
XX

DE Human ovarian tumour associated polynucleotide sequence SEQ ID NO:1429.  
XX  
KW Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;  
KW immunogenic; vaccine; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200151513-A2.  
XX  
PD 19-JUL-2001.  
XX  
PE 16-JAN-2001; 2001WO-US01575.  
XX  
PR 14-JAN-2000; 2000US-0176722.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Algate PA;  
XX  
DR WPI: 2001-425866/45.  
XX  
PT Novel ovarian tumor proteins, and nucleic acids encoding them, used to  
PT treat and diagnose cancers, particularly ovarian cancer -  
XX  
PS Claim 5; Page 325; 338pp; English.  
XX  
CC AAH82377 to AAH83878 represent human ovarian tumour-associated  
CC polynucleotide sequences which encode ovarian tumour proteins. The  
CC ovarian tumour protein and polynucleotide sequences have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. The  
CC ovarian tumour proteins and polynucleotides can be used to inhibit  
CC the development of cancer, particularly ovarian cancer. They can also  
CC be used to diagnose the onset and progression of cancer.  
XX  
SQ Sequence 346 BP; 90 A; 64 C; 66 G; 126 T; 0 other;

Query Match 78.1%; Score 16.4; DB 22; Length 346;  
Best Local Similarity 94.4%; Pred. No. 3.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaa 18  
Db 25 atgaaagcaagctaa 42

RESULT 15  
AAH83806  
ID AAH83806 standard; cDNA: 346 BP.  
XX  
AC AAH83806;  
XX  
DT 25-SEP-2001 (first entry)  
XX  
DE Human ovarian tumour associated polynucleotide sequence SEQ ID NO:1430.  
XX  
KW Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;  
KW immunogenic; vaccine; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200151513-A2.  
XX  
PD 19-JUL-2001.  
XX  
PE 16-JAN-2001; 2001WO-US01575.  
XX  
PR 14-JAN-2000; 2000US-0176722.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Algate PA;  
XX



DR WPI: 2001-425866/45.

XX  
PT Novel ovarian tumor proteins, and nucleic acids encoding them, used to  
PT treat and diagnose cancers, particularly ovarian cancer .  
XX  
XX  
PS Claim 5; Page 325; 338pp; English.

XX  
CC AAH82377 to AAH83878 represent human ovarian tumour-associated  
CC polynucleotide sequences which encode ovarian tumour proteins. The  
CC ovarian tumour protein and polynucleotide sequences have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. The  
CC ovarian tumour proteins and polynucleotides can be used to inhibit  
CC the development of cancer, particularly ovarian cancer. They can also  
CC be used to diagnose the onset and progression of cancer.

XX  
SQ Sequence 346 BP; 89 A; 64 C; 66 G; 126 T; 1 other;

Query Match 78.1%; Score 16.4; DB 22; Length 346;  
Best Local Similarity 94.4%; Pred. No. 3.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 atgaaaggaagctaaa 18  
||||||| |||||||  
Db 25 atgaaagcaagctaaa 42

Search completed: March 20, 2002, 01:57:37  
Job time: 7018 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2002, 00:45:13 ; Search time 2311.81 seconds  
(without alignments)  
97.612 Million cell updates/sec

Title: US-09-004-395-3

Perfect score: 21

Sequence: 1 atgaagaaggaagctaaagt 21

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estcom:\*  
5: em\_estcpl:\*  
6: em\_estlba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gb\_estl:\*  
11: gb\_estl2:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	92.4	857	13	CNSO1NND
2	19.4	92.4	863	13	CNSO1NUR
3	18.4	87.6	211	10	AA514077
4	18.4	87.6	225	10	AV276062
5	18.4	87.6	449	11	BG653289
6	18.4	87.6	457	13	AQ151824
7	18.4	87.6	654	13	AZ992563
8	18.4	87.6	717	10	BE537306
9	18.4	87.6	887	10	BE539217
10	18	85.7	281	10	BB326258
11	18	85.7	308	10	BB552761
12	18	85.7	467	13	AZ936827

13	17.8	84.8	217	10	AA700025	AA700025 2169412.s
14	17.8	84.8	296	11	N73600	N73600 2a51d06.s1
15	17.8	84.8	311	10	AV219438	AV219438 AV219438
16	17.8	84.8	321	10	AL586792	AL586792 AL586792
17	17.8	84.8	324	10	BB494627	BB494627 BB494627
18	17.8	84.8	355	11	D63016	D63016 HUM352D07B
19	17.8	84.8	362	13	AQ824834	AQ824834 HS_3320.B
20	17.8	84.8	367	13	AO146623	AO146623 HS_2233.B
21	17.8	84.8	397	11	T87926	T87926 ye08c06.r1
22	17.8	84.8	401	11	D62534	D62534 HUM300B06B
23	17.8	84.8	414	10	AA987196	AA987196 Or92a11.s
24	17.8	84.8	427	13	AZ666735	AZ666735 1M0549003
25	17.8	84.8	428	11	R19941	R19941 Y930F02.r1
26	17.8	84.8	442	10	A1694973	A1694973 w444h08.x
27	17.8	84.8	473	11	BG055061	BG055061 nac94h08.
28	17.8	84.8	478	10	A1953318	A1953318 wq27a01.x
29	17.8	84.8	491	10	A1371083	A1371083 tA07f09.x
30	17.8	84.8	506	10	A1589168	A1589168 tF81h04.x
31	17.8	84.8	508	10	AM662149	AM662149 h120b12.x
32	17.8	84.8	511	10	A1478812	A1478812 tm52f03.x
33	17.8	84.8	518	13	AZ479417	AZ479417 1M0300G01
34	17.8	84.8	519	11	BF447906	BF447906 7q96h10.x
35	17.8	84.8	579	10	BE669553	BE669553 7e15a06.x
36	17.8	84.8	592	10	A1179333	A1179333 EST223029
37	17.8	84.8	598	10	AM077421	AM077421 f336a07.x
38	17.8	84.8	677	11	BG292236	BG292236 602386571
39	17.8	84.8	711	10	AV716967	AV716967 AV716967
40	17.8	84.8	738	11	BG611921	BG611921 602613364
41	17.8	84.8	754	11	BG283645	BG283645 602407992
42	17.8	84.8	759	11	BF683582	BF683582 963045807
43	17.8	84.8	872	13	AZ186410	AZ186410 SP_1006.B
44	17.8	84.8	991	13	CNS0784T	AL434043 T3 end of
45	17.4	82.9	296	10	BB229099	BB229099 BB229099

#### ALIGNMENTS

RESULT 1	CNSO1NND	857 bp	DNA	GSS	14-JUN-2001
LOCUS	CNSO1NND	857 bp	DNA		
DEFINITION	Anopheles gambiae GSS sp6 end of clone 24H10 of NotreDame1 library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.				
ACCESSION	AL152442	GI:7013361			
VERSION	AL152442.1				
KEYWORDS	GSS.				
SOURCE	African malaria mosquito.				
ORGANISM	Anopheles gambiae				
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.				
AUTHORS	1 (bases 1 to 857)				
TITLE	Genoscope.				
JOURNAL	Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
REFERENCE	2 (bases 1 to 857)				
AUTHORS	Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weisszenbach,J.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-FEB-2000) BBMT, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France				
COMMENT	This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.				
FEATURES	Location/Qualifiers				
source	1..857				
	/organism="Anopheles gambiae"				
	/strain="PEST"				
	/db_xref="taxon:7165"				
	/clone="24H10"				

```

/clone.lib="Notredame1"
/note="end : SP6"
BASE COUNT      270 a      151 c      142 g      290 t      4 others
ORIGIN

Query Match      92.4%: Score 19.4; DB 13; Length 857;
Best Local Similarity 95.2%: Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 atgaaaggaagctaaagt 21
|||||
Db 601 ATGAAAGGAAGCTAAAGT 621

RESULT 2
CONS10NR      863 bp      DNA      GSS      14-JUN-2001
LOCUS      Anopheles gambiae GSS SP6 end of clone 24005 of Notredame1 library
DEFINITION      from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION      AL152708
VERSION      AL152708.1 GI:7013627
KEYWORDS      African malaria mosquito.
SOURCE      Anopheles gambiae
ORGANISM      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Anopheles.
REFERENCE      1 (bases 1 to 863)
AUTHORS      Genoscope.
TITLE      Direct Submision
JOURNAL      Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : segrefigenoscope.cns.fr
Web : www.genoscope.cns.fr)
2 (bases 1 to 863)
Roch,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
Direct Submision
Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES
Source      1..863
Location/Qualifiers
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="24005"
/clone.lib="Notredame1"
/note="end : SP6"
BASE COUNT      275 a      153 c      142 g      291 t      2 others
ORIGIN

Query Match      92.4%: Score 19.4; DB 13; Length 863;
Best Local Similarity 95.2%: Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 atgaaaggaagctaaagt 21
|||||
Db 578 ATGAAAGGAAGCTAAAGT 598

RESULT 3
AA514077      211 bp      mRNA      EST      10-JUL-1997
LOCUS      MC2AS032.ACS S. mansoni cercarial lambda zap Schistosoma mansoni
DEFINITION      CDNA clone AS-32 5', mRNA sequence.
ACCESSION      AA514077
VERSION      AA514077.1 GI:2252933
KEYWORDS      EST.

```

```

SOURCE      Schistosoma mansoni.
ORGANISM      Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
Rhabdiphora; Eulicthophora; Revertepermatata; Medifusata;
Neodermata; Trematoda; Digenea; Strigilidae; Schistosomatidae;
Schistosoma.
REFERENCE      1 (bases 1 to 211)
AUTHORS      Karim,A.M.
TITLE      Schistosome Gene Identification
JOURNAL      Unpublished (1997)
COMMENT      Contact: Karim,A.M.
Dr.Amr Karim, Department of Biochemistry
Faculty of Science, Ain Shams University
Box 11456, El-Khalifa El-Maamoun St., Abbasia, Cairo, Egypt
Tel: (202)2853561
Fax: (202)2853561
Email: Hadyeasunet.shams.eun.eg
Seq primer: SK
High quality sequence stop: 211.
FEATURES
Source      1..211
Location/Qualifiers
/organism="Schistosoma mansoni"
/db_xref="taxon:6183"
/clone="AS-32"
/clone.lib="S. mansoni cercarial lambda zap"
/sex="Mixed"
/dev_stage="cercarial"
/note="library provided by Dr. M. Saber TBRI,Cairo,Egypt."
BASE COUNT      61 a      36 c      51 g      63 t
ORIGIN

Query Match      87.6%: Score 18.4; DB 10; Length 211;
Best Local Similarity 95.0%: Pred. No. 3.8e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 tgaaggaagctaaagt 21
|||||
Db 153 TGAAGGAAGCTAAAGT 172

RESULT 4
AV276062      225 bp      mRNA      EST      05-NOV-1999
LOCUS      AV276062 RIKEN full-length enriched, adult male testis (DH10B) Mus
DEFINITION      Musculus CDNA clone 4932432H20 3', mRNA sequence.
ACCESSION      AV276062
VERSION      AV276062.1 GI:6264099
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 225)
AUTHORS      Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Isnii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Suganara,Y., Suzuki,H.,
Suzuki,H., Takahashi,F., Tateo,M., Tomihaga,N., Tsunoda,Y.,
Wataniki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,Y.,
Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

```



JOURNAL  
MEDLINE  
COMMENT

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3080 row: P column: 21  
Class: BAC ends  
High quality sequence stop: 457.  
Location/Qualifiers

FEATURES  
source

1. .457  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate3080 Col-21 Row-P"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC1; BAC clones in E-Coli DH10B"

BASE COUNT 135 a 92 c 72 g 158 t

ORIGIN

Query Match 87.6%; Score 18.4; DB 13; Length 457;  
Best Local Similarity 95.0%; Pred. No. 3.1e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaaag 20  
|||||  
Db 163 ATGAACAGCAAGCTAAAG 144

RESULT 7  
AZ992563

LOCUS 654 bp DNA GSS 27-APR-2001  
DEFINITION 2M0277G14F Mouse 10kb plasmid UNGC2M library Mus musculus genomic  
clone UNGC2M0277G14 F, DNA sequence.

ACCESSION AZ992563  
VERSION AZ992563.1 GI:13863790

KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.,  
and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0277 row: G column: 14  
Seq Primer: CGTGTGAACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 654.  
Location/Qualifiers

FEATURES  
source

1. .654  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"

/clone="UNG2M0277G14"  
/clone\_lib="Mouse 10kb plasmid UNGC2M library"  
/sex="Female"  
/lab\_host="E. coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g11473211419b1AF129072.1) a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 215 a 113 c 138 g 188 t

ORIGIN

Query Match 87.6%; Score 18.4; DB 13; Length 654;  
Best Local Similarity 95.0%; Pred. No. 2.8e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaaag 20  
|||||  
Db 557 ATGAACAGCAAGCTAAAG 576

RESULT 8  
BE537306

LOCUS 717 bp mRNA EST 09-AUG-2000  
DEFINITION 601065480P1 NIH\_MGC\_10 Homo sapiens cDNA clone IMAGE:3452011 5',  
mRNA sequence.

ACCESSION BE537306  
VERSION BE537306.1 GI:9765951

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Incyte Genomics, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L1AM8432 row: 0 column: 20  
High quality sequence stop: 678.  
Location/Qualifiers

FEATURES  
source

1. .717  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3452011"  
/clone\_lib="NIH\_MGC\_10"  
/cell\_line="MGC36"  
/lab\_host="DH10B"  
/note="Organ: cervix; Vector: PCMV-SPO6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.



BASE COUNT	57 a	67 c	40 g	116 t	1 others									
ORIGIN														
Query Match	85.7%; Score 18; DB 10; Length 281;													
Best Local Similarity	100.0%; Pred. No. 4.8e+03;													
Matches	18; Conservative	0; Mismatches	0; Indels	0; Gaps	0;									
Qy	3	gaaaggaagcttaag	20											
Db	170	GAAAGGAAAGCTTAAG	153											
RESULT 11														
LOCUS	BB552761/c													
DEFINITION	BB552761 RIKEN full-length enriched, 2 days pregnant adult female	EST	01-AUG-2000											
ACCESSION	BB552761													
VERSION	BB552761.1	GI:9639127												
KEYWORDS	EST.													
SOURCE	house mouse.													
ORGANISM	Mus musculus													
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus. 1 (bases 1 to 308)													
AUTHORS	Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koyu, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shimomoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugihara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamataka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.													
TITLE	RIKEN Mouse ESTs (Konno, H., et al.)													
JOURNAL	Unpublished (2000)													
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasakihara, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Thermosensitization and thermoinactivation of thermolabile enzymes by thermosensitization and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, K., Ozawa, T., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site (http://genome.ritc.riken.go.jp) for further details. Location/Qualifiers 1. 308 /organism="Mus musculus" /db_xref="taxon:10090" /clone_id="E330001H21" /clone_lib="RIKEN full-length enriched, 2 days pregnant adult female ovary" /sex="female" /tissue_type="ovary"													

```

/def_stage="2 days pregnant adult"
/lab_host="DH10B"
/note="Site_1: Sali. Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGGAGAGAGCGCGCCGACACCTGAGCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGGAGAGAGCTTCTCGATTATTAATTAATGCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT      70 a      75 c      35 g      128 t

ORIGIN

Query Match      85.7%; Score 18; DB 10; Length 308;
Best Local Similarity 100.0%; Pired. No. 4.7e+03;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 gaaaggaagaaactaaag 20
        |||||||
Db      138 GAAAGGAAAGCTTAAG 121

RESULT 12
A2936827      467 bp      DNA      GSS      26-APR-2001
LOCUS      2M0193D13R Mouse 10kb plasmid UNGC2M library Mus musculus genomic
DEFINITION      clone UNGC2M0193D13 R, DNA sequence.
ACCESSION      A2936827
VERSION
KEYWORDS
SOURCE
ORGANISM      house mouse.
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 467)
REFERENCE      Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C.,
AUTHORS      Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly
            ,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A.
            and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
TITLE      unpublished (2000)
JOURNAL
COMMENT      Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
            84112 USA
            Tel.: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0193 row: D column: 13
            Seq primer: CACACAGAGAAACACCTATGACC
            Class: plasmid ends
            High quality sequence stop: 467.
            Location/Qualifiers
                1..467
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="UNG2M0193D13"
                    /clone_1db="Mouse 10kb plasmid UNGC2M library"
                    /sex="Female"
                    /lab_host="E. coli strain XL10-Gold, TI-resistant, F-"
                    /note="Vector: pMD42nv; Purified genomic DNA from M.

```



musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 92 a 111 c 146 g 115 t 3 others  
ORIGIN

Query Match 85.7%; Score 18; DB 13; Length 467;  
Best Local Similarity 100.0%; Pred. No. 4.2e+03;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 aaaaagaaagctaaagt 21  
|||||  
Db 34 AAAAGGAAAGCTAAAGT 51

RESULT 13  
AA700025 217 bp mRNA EST 19-DEC-1997  
LOCUS z169d12.s1 Soares\_fetal\_liver\_spleen\_INFLS\_S1 Homo sapiens CDNA  
DEFINITION clone IMAGE:436055 3', mRNA sequence.  
ACCESSION AA700025  
VERSION AA700025.1 GI:2702988  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 217)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Getzel, G., Jost, S.,  
Kizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,  
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 107.

## FEATURES

source  
1..217  
/organism="Homo sapiens"  
/db\_xref="GDB:1335824"  
/db\_xref="taxon:9606"  
/clone="IMAGE:436055"  
/clone\_1lb="Soares\_fetal\_liver\_spleen\_INFLS\_S1"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and spleen; Vector: p773D (Pharmacia)  
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;  
This is a subtracted version of the original Soares fetal

liver spleen INFLS library. 1st strand cDNA was primed  
with a Pac I - oligo(dT) primer [5',  
AACGCGAAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified p773 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 72 a 47 c 54 g 44 t  
ORIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 217;  
Best Local Similarity 90.5%; Pred. No. 6.1e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atgaagaaagctaaagt 21  
|||  
Db 50 ATGTAAGGAAGCTAAAT 70

RESULT 14  
N73600 296 bp mRNA EST 19-MAR-1996  
LOCUS z451d06.s1 Soares\_fetal\_liver\_spleen\_INFLS\_Homo sapiens CDNA clone  
DEFINITION IMAGE:296075 3' similar to contains Alu repetitive element; mRNA  
sequence.  
ACCESSION N73600  
VERSION N73600.1 GI:1230885  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 296)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,  
R., Williamson, A., Wohldmann, P. and Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
Seq primer: m13 -40 forward  
High quality sequence stop: 212.

## FEATURES

source  
1..296  
/organism="Homo sapiens"  
/db\_xref="GDB:1240966"  
/db\_xref="taxon:9606"  
/clone="IMAGE:296075"  
/clone\_1lb="Soares fetal liver spleen INFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)  
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' AACGCGAAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified p773 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 95 a 64 c 68 g 66 t 3 others  
ORIGIN







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2002, 01:30:37 ; Search time 116.86 Seconds

(without alignments)  
40.699 Million cell updates/sec

Title: US-09-004-395-3

Perfect score: 21  
Sequence: 1 atgaaagaagactaaagt 21

Scoring table: IDENTITY\_NUC  
Gap 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PCITUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	16.2	77.1	594	1	US-08-404-732A-3
C 2	16.2	77.1	1420	1	US-08-080-255-3
C 3	16.2	77.1	1420	3	US-08-465-713-3
C 4	16.2	77.1	1420	5	PCIT-US93-05857-3
C 5	16.2	77.1	1853	1	US-08-404-732A-6
C 6	16.2	77.1	2108	4	US-09-032-742-6
C 7	16.2	77.1	2120	1	US-08-404-732A-4
C 8	16.2	77.1	2224	1	US-08-404-732A-8
C 9	16.2	77.1	2325	2	US-08-966-388-5
C 10	16.2	77.1	2325	3	US-09-188-403-5
C 11	16.2	77.1	2325	3	US-09-188-403-5
C 12	16.2	77.1	2325	3	US-09-188-403-5
C 13	16.2	77.1	2325	4	US-09-281-259-5
C 14	16.2	77.1	3956	2	US-08-966-388-9
C 15	16.2	77.1	3956	3	US-09-188-403-9
C 16	16.2	77.1	3956	3	US-09-188-404-9
C 17	16.2	77.1	3956	4	US-09-281-259-9
C 18	16.2	77.1	4201	1	US-08-080-255-4
C 19	16.2	77.1	4201	1	US-08-465-713-4
C 20	16.2	77.1	4201	5	PCIT-US93-05857-4
C 21	16.2	77.1	11907	4	US-08-061-376-4
C 22	16.2	77.1	14255	1	US-08-320-559-1
C 23	16.2	77.1	14255	1	US-08-327-392-1
C 24	16.2	77.1	14255	1	US-08-306-691B-55
C 25	16.2	77.1	14255	3	US-08-545-860D-1
C 26	16.2	77.1	14255	5	PCIT-US94-04496-1
C 27	15.8	75.2	414	4	US-09-330-330-3
					Sequence 33, Appl

28	15.8	75.2	1012	4	US-09-446-504-32	Sequence 32, Appl
29	15.8	75.2	1554	2	US-08-031-538-8	Sequence 8, Appl
30	15.8	75.2	1730	1	US-07-817-920-1	Sequence 1, Appl
31	15.8	75.2	1730	1	US-08-117-006-1	Sequence 1, Appl
32	15.8	75.2	1730	1	US-08-216-594-1	Sequence 1, Appl
33	15.8	75.2	1730	5	PCIT-US93-00149-1	Sequence 1, Appl
34	15.8	75.2	2781	3	US-08-749-522-4	Sequence 4, Appl
35	15.4	73.3	4431	4	US-09-532-803-8	Sequence 8, Appl
36	15.4	73.3	12145	3	US-08-968-563-19	Sequence 19, Appl
37	15.4	73.3	12145	3	US-08-969-683A-19	Sequence 19, Appl
38	15.2	72.4	913	1	US-08-109-391A-1	Sequence 1, Appl
39	15.2	72.4	913	1	US-08-459-019A-1	Sequence 1, Appl
40	15.2	72.4	913	2	US-08-460-428A-1	Sequence 1, Appl
41	15.2	72.4	913	3	US-08-458-860A-1	Sequence 1, Appl
42	15.2	72.4	1972	1	US-08-463-048-1	Sequence 1, Appl
43	15.2	72.4	1972	1	US-08-463-229-1	Sequence 1, Appl
44	15.2	72.4	1972	2	US-08-302-891-1	Sequence 1, Appl
45	15.2	72.4	3763	1	US-07-792-865D-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-404-732A-3/C  
Sequence 3, Application US/08404732A  
Patent No. 5661017  
GENERAL INFORMATION:  
APPLICANT: Dunahay, Terry G.  
APPLICANT: Roesler, Paul  
APPLICANT: Jarvis, Eric  
TITLE OF INVENTION: METHOD TO TRANSFORM ALGAE, MATERIALS  
TITLE OF INVENTION: THEREFOR, AND PRODUCTS PRODUCED THEREBY  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: National Renewable Energy Laboratory  
STREET: 1617 Cole Boulevard  
CITY: Golden  
STATE: CO  
COUNTRY: U.S.A.  
ZIP: 80401-3393  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(DOS)text (\*.\*)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/404,732A  
FILING DATE: 15-03-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: O'Connor, Edna M.  
REGISTRATION NUMBER: 29252  
REFERENCE/DOCKET NUMBER: 94-31  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/275-4400  
TELEFAX: 303/275-4400  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 594 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 1..594  
US-08-404-732A-3

Query Match 77.1%; Score 16.2; DB 1; Length 594;  
Best Local Similarity 85.7%; Pred. No. 95;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 atgaaaggaaagctaaagt 21  
||||| || |||||  
DB 100 ATGAAGAGCAATCCTAAAGT 80

## RESULT 2

US-08-080-255-3/c  
; Sequence 3, Application US/08080255  
; Patent No. 5487970  
; GENERAL INFORMATION:  
; APPLICANT: Rowley, Janet D.  
; APPLICANT: Diaz, Manuel O.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND  
; TITLE OF INVENTION: TRANSLOCATIONS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/080.255  
; FILING DATE: 19930617  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165  
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 320-7200  
; TELEFAX: (512) 474-7577  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1420 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-080-255-3

Query Match 77.1%; Score 16.2; DB 1; Length 1420;  
Best Local Similarity 85.7%; Pred. No. 98;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 atgaaaggaaagctaaagt 21  
||||| ||||| |||||  
DB 70 ATGAAGAGCAAAAGTAAAGT 50

## RESULT 3

US-08-465-713-3/c  
; Sequence 3, Application US/08465713  
; Patent No. 6121419  
; GENERAL INFORMATION:  
; APPLICANT: Rowley, Janet D.  
; APPLICANT: Diaz, Manuel O.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND  
; TITLE OF INVENTION: TRANSLOCATIONS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433

CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465.713  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/080.255  
FILING DATE: 17 JUNE 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: ARCD:072/PAR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 320-7200  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1420 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-465-713-3

Query Match 77.1%; Score 16.2; DB 3; Length 1420;  
Best Local Similarity 85.7%; Pred. No. 98;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 atgaaaggaaagctaaagt 21  
||||| ||||| |||||  
DB 70 ATGAAGAGCAAAAGTAAAGT 50

## RESULT 4

PCT-US93-05857-3/c  
; Sequence 3, Application PC/TUS9305857  
; GENERAL INFORMATION:  
; APPLICANT: Board of Regents  
; APPLICANT: The University of Texas System  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING  
; TITLE OF INVENTION: GENE REARRANGEMENTS AND TRANSLOCATIONS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/05857  
; FILING DATE: 19930617  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/900,689  
; FILING DATE: 17/06/92  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165

REFERENCE/DOCKET NUMBER: ARCD:072/PAR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 320-7200  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1420 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
PCT-US93-05857-3

Query Match 77.1% Score 16.2; DB 5; Length 1420;  
Best Local Similarity 85.7%; Pred. No. 98;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 atgaagaagaaagctaaagt 21  
|||||  
Db 70 ATGAGAGCAAGCAATCTAAAGT 50

RESULT 5  
US-08-404-732A-6/c  
Sequence 6, Application US/08404732A  
Patent No. 5661017  
GENERAL INFORMATION:  
APPLICANT: Dunahay, Terry G.  
APPLICANT: Roesler, Paul  
APPLICANT: Jarvis, Eric  
TITLE OF INVENTION: METHOD TO TRANSFORM ALGAE MATERIALS  
TITLE OF INVENTION: THEREFOR, AND PRODUCTS PRODUCED THEREBY  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: National Renewable Energy Laboratory  
STREET: 1617 Cole Boulevard  
CITY: Golden  
STATE: CO  
COUNTRY: U.S.A.  
ZIP: 80401-3393  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(DOS)text (\*.\*)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/404,732A  
FILING DATE: 15-03-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: O'Connor, Edna M.  
REGISTRATION NUMBER: 29252  
REFERENCE/DOCKET NUMBER: 94-31  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/275-4404  
TELEFAX: 303/275-4400  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1853 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: 1, 5'UTR  
LOCATION: 1, 445  
OTHER INFORMATION: /label- Accase  
NAME/KEY: CDS  
LOCATION: 446..1240  
OTHER INFORMATION: /label- NPTII  
NAME/KEY: Linker Sequence

LOCATION: 1241..1243  
NAME/KEY: Accase 3' coding region (in non-translatable reading  
frame)  
NAME/KEY: 1244..1259  
LOCATION: 1244..1259  
NAME/KEY: 3'UTR  
LOCATION: 1260..1853  
OTHER INFORMATION: /label- Accase  
US-08-404-732A-6

Query Match 77.1% Score 16.2; DB 1; Length 1853;  
Best Local Similarity 85.7%; Pred. No. 99;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 atgaagaagaaagctaaagt 21  
|||||  
Db 1359 ATGAAAGCAATCTAAAGT 1339

RESULT 6  
US-09-032-742-6/c  
Sequence 6, Application US/09032742  
Patent No. 6255089  
GENERAL INFORMATION:  
APPLICANT: Teitler, Milt  
APPLICANT: Herrick-Davis, Katharine  
APPLICANT: Egan, Christina C.  
TITLE OF INVENTION: Constitutively Activated Serotonin  
TITLE OF INVENTION: Receptors  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Laurence Weinberger  
STREET: 882 S. MacLack Street, Suite 103  
CITY: P.O. Box 1663  
STATE: PA  
COUNTRY: USA  
ZIP: 19380-0053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/032,742  
FILING DATE: 27-FEB-1998  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Weinberger, Laurence  
REGISTRATION NUMBER: 27,965  
REFERENCE/DOCKET NUMBER: 3086-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 431-1703  
TELEFAX: (610) 431-4181  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2108 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-032-742-6

Query Match 77.1% Score 16.2; DB 4; Length 2108;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 atgaagaagaaagctaaagt 21  
|||||  
Db 1575 ATGAAAGCAATCTAAAGT 1555

```
RESULT 7
US-08-404-732A-4/C
: Sequence 4, Application US/08404732A
: Patient No. 5661017
: GENERAL INFORMATION:
: APPLICANT: Dunahay, Terry G.
: APPLICANT: Roesler, Paul
: APPLICANT: Jarvis, Eric
: TITLE OF INVENTION: METHOD TO TRANSFORM ALGAE, MATERIALS
: THEREFOR, AND PRODUCTS PRODUCED THEREBY
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: National Renewable Energy Laboratory
: STREET: 1617 Cole Boulevard
: CITY: Golden
: STATE: CO
: COUNTRY: U.S.A.
: ZIP: 80401-3393
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII(DOS)text (*.*)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/404,732A
: FILING DATE: 15-030-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: O'Connor, Edna M.
: REGISTRATION NUMBER: 29252
: REFERENCE/DOCKET NUMBER: 94-31
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 303/275-4400
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2120 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: 1.
: NAME/KEY: 5'UTR
: LOCATION: 1..445
: OTHER INFORMATION: /label= Accase
: NAME/KEY: CDS
: LOCATION: 446..1240
: OTHER INFORMATION: /label= NPITI
: NAME/KEY: Linker Sequence
: LOCATION: 1241..1248
: NAME/KEY: ACCase 3' coding region (in non-translatable reading
: frame)
: LOCATION: 1249..1526
: NAME/KEY: 3'UTR
: LOCATION: 1527..2120
: OTHER INFORMATION: /label= Accase
: US-08-404-732A-4

Query Match 77.1%; Score 16.2; DB 1; Length 2120;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaaagt 21
||||| ||| ||| |||
Db 1626 ATGAAAGCAATCCTAAAGT 1606

RESULT 8
US-08-404-732A-8/C
: Sequence 8, Application US/08404732A
: Patient No. 5661017
```

```
: GENERAL INFORMATION:
: APPLICANT: Dunahay, Terry G.
: APPLICANT: Roesler, Paul
: APPLICANT: Jarvis, Eric
: TITLE OF INVENTION: METHOD TO TRANSFORM ALGAE, MATERIALS
: THEREFOR, AND PRODUCTS PRODUCED THEREBY
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: National Renewable Energy Laboratory
: STREET: 1617 Cole Boulevard
: CITY: Golden
: STATE: CO
: COUNTRY: U.S.A.
: ZIP: 80401-3393
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII(DOS)text (*.*)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/404,732A
: FILING DATE: 15-030-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: O'Connor, Edna M.
: REGISTRATION NUMBER: 29252
: REFERENCE/DOCKET NUMBER: 94-31
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 303/275-4400
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2224 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: 1.
: NAME/KEY: 5'UTR
: LOCATION: 1..816
: OTHER INFORMATION: /label= Accase
: NAME/KEY: CDS
: LOCATION: 817..1611
: OTHER INFORMATION: /label= NPITI
: NAME/KEY: Linker Sequence
: LOCATION: 1612..1614
: NAME/KEY: ACCase 3' coding region (in non-translatable reading
: frame)
: LOCATION: 1615..1630
: NAME/KEY: 3'UTR
: LOCATION: 1631..2224
: OTHER INFORMATION: /label= Accase
: US-08-404-732A-8

Query Match 77.1%; Score 16.2; DB 1; Length 2224;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaaagt 21
||||| ||| ||| |||
Db 1730 ATGAAAGCAATCCTAAAGT 1710

RESULT 9
US-08-966-388-5
: Sequence 5, Application US/08966388
: Patient No. 5965412
: GENERAL INFORMATION:
: APPLICANT: TOMOYUKI NISHIMOTO
: APPLICANT: MICHIO KUBOTA
: APPLICANT: HIROTO CHAEN
```



```

1  APPLICANT: TOSHIO MIYAKE
2  TITLE OF INVENTION: KUIBILOSE PHOSPHORYLASE, ITS PREPARATION AND USES
3  NUMBER OF SEQUENCES: 10
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: BROWDY AND NEIMARK
6  STREET: 419 Seventh Street, N.W., Suite 300
7  CITY: Washington
8  STATE: D.C.
9  COUNTRY: USA
10 ZIP: 20004
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: Patent in Release #1.0, Version #1.30
17
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/966,388
20 FILING DATE:
21 CLASSIFICATION: 435
22
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: JP 311,235/1996
25 FILING DATE: 8-NOV-1996
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: JP 61,710/97
28 FILING DATE: 3-MAR-1997
29 ATTORNEY/AGENT INFORMATION:
30 NAME: BROWDY, Roger L.
31 REGISTRATION NUMBER: 25,618
32 REFERENCE/DOCKET NUMBER:
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE: 202-628-5197
35 TELEFAX: 202-737-3528
36
37 TELEX: 248633
38
39 INFORMATION FOR SEQ ID NO: 5:
40
41 SEQUENCE CHARACTERISTICS:
42 LENGTH: 2325 base pairs
43 TYPE: nucleic acid
44 STRANDEDNESS: double
45 TOPOLOGY: linear
46
47 US-08-966-388-5
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
```

```

SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/188,403
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/966,388
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 61,710/97
FILING DATE: 3-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2325 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-188-403-5

Query Match 77.1%; Score 16.2; DB 3; Length 2325;
Best Local Similarity 85.7%; Pred. No.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0.

QY 1 atgaaagaagaaagctaaagt 21
   1 ||||| 111111 |||||
Db 2274 AGGAAAGCAATATATAAGT 2294

RESULT 11
US-09-188-404-5
Sequence 5, Application US/09188404
Patent No. 6140487
GENERAL INFORMATION:
APPLICANT: TOMOYUKI NISHIMOTO
APPLICANT: MICHIO KUBOTA
APPLICANT: HIROO CHAEN
APPLICANT: TOSHIKO MIYAKE
TITLE OF INVENTION: KOTIBIOSE PHOSPHORYLASE, ITS PREPARATION AND USES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/188,404
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/966,388
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 61,710/97
FILING DATE: 3-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618

```

REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2325 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-188-404-5

Query Match 77.1%; Score 16.2; DB 3; Length 2325;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgaagaaggaagctaaagt 21  
Db 2274 AGGAAAGGAATATATAAGT 2294

RESULT 12  
US-09-281-259-5  
Sequence 5, Application US/09281259  
Patent No. 6204377  
GENERAL INFORMATION:  
APPLICANT: TOMOYUKI NISHIMOTO  
APPLICANT: MICHIO KUBOTA  
APPLICANT: HIROTO CHAEN  
APPLICANT: TOSHIO MIYAKE  
TITLE OF INVENTION: KOJIBIOSE PHOSPHORYLASE, ITS PREPARATION AND USES  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/281,259  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/966,388  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 61,710/97  
FILING DATE: 3-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2325 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-281-259-5

Query Match 77.1%; Score 16.2; DB 4; Length 2325;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgaagaaggaagctaaagt 21  
Db 2274 AGGAAAGGAATATATAAGT 2294

RESULT 13  
US-08-966-388-9  
Sequence 9, Application US/08966388  
Patent No. 5965412  
GENERAL INFORMATION:  
APPLICANT: TOMOYUKI NISHIMOTO  
APPLICANT: MICHIO KUBOTA  
APPLICANT: HIROTO CHAEN  
APPLICANT: TOSHIO MIYAKE  
TITLE OF INVENTION: KOJIBIOSE PHOSPHORYLASE, ITS PREPARATION AND USES  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/966,388  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 311,235/1996  
FILING DATE: 8-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 61,710/97  
FILING DATE: 3-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3956 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Thermoaerobium brockii  
STRAIN: ATCC 35047  
FEATURE:  
FEATURE: 1-100 5'-UTR  
FEATURE: E  
FEATURE: 101-2425 mat peptide  
FEATURE: S  
FEATURE: 2426-3956 3'-UTR  
FEATURE: E  
US-08-966-388-9

Query Match 77.1%; Score 16.2; DB 2; Length 3956;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



Search completed: March 20, 2002, 01:30:39  
Job time: 6505 sec

---



Db 226 IVPYRSPAPLKLKGIKIHRSOEGDVSIRKIDKIVYDQAVVNRNSVDDEALMCI 285  
OY 285 -----FKVYETSGTESLRKKAHETEK 306  
Db 286 LKREBOYRNFELAKLGNLQVLRSLKKK 314

## RESULT 2

O9LA17 PRELIMINARY: PRT: 285 AA.

AC O9LA17: 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DR 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE flaa-1 (FRAGMENT).  
CN  
OS Spirochete str. PFRS-A.  
OC Bacteria; Spirochaetales; Spirochaetaceae.  
OX NCBI\_TaxID=120784;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=PFRS-A.  
RA Sato T., Kuramitsu H.K.;  
RT "Polymerase chain reaction for the detection of flaa-1 genes of oral  
RT Spirochetes in advanced periodontal pockets."  
RL Arch. Oral Biol. 0:0-0(2000).  
DR EMBL: AF132816; AAF64553.1; -  
FT NON-TER 1  
FT SEQUENCE 285 AA: 31864 MM: 4D7FCC4BF5483C9F CRC64:

Query Match 20.0%; Score 343; DB 2; Length 285;  
Best Local Similarity 29.4%; Pred. No. 2,7e-16;  
Matches 90; Conservative 56; Mismatches 114; Indels 46; Gaps 8;

OY 13 ISTDVFAOETDGLAGSKRAEPGLVDFAEIARDPSSSTRDLDTNVDVYSGAGSIYKP 72  
Db 2 LNADIMADKSGGTMNRR-----TVLDYASLAD-----TSYTD-----EQK 37  
OY 73 EDMVVDLGINNMSVLLTPSARLQAVYKNSVAPAVYKSKSKRYAGTILGVRVLPSPY-S 131  
Db 38 ALMRSSLAQAQVEVNLNSASRPVAHAASRVLEAPVSEGAISFAGRGVLRVLEPTWDS 97  
OY 133 OSSAMIMPFKIPFYK-----GEGN-----QPL-GKGLIDNKTKEIKY 171  
Db 98 NANNMIKPAFVIPAYEVMAQVDDQGNVQAPTEEEKASGKRGREDGIGVYKNGVLAKSTAV 157  
OY 172 SVYSLGVEIDLEVLFDNMGMEYASMGTLKFKGNADLIWSPNYIPNISRIKDDVFN 231  
Db 158 NTYGMNYPHGLVMMRDQGEVHRFYMGYLLFPDSWKELVNPNPSYISDVRSREVR-LYPV 216  
OY 232 VPLASSKMFKAFRVSKSSKSKVKNFIFYVKDLRVLYDKLSISDSDISESVFYKYEYS 291  
Db 217 YASPPHVVVEGFVWTRDAAHAGDYGVGKDKIITYDKAVLSTVRDEADDELWGIQARR 276  
OY 292 GTESLR 297  
Db 277 EAERRK 282

## RESULT 3

O9LA16 PRELIMINARY: PRT: 285 AA.

AC O9LA16: 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DR 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE flaa-1 (FRAGMENT).  
CN  
OS Spirochete str. PFRS-B.  
OC Bacteria; Spirochaetales; Spirochaetaceae.  
OX NCBI\_TaxID=120785;

RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=PFRS-B.  
RA Sato T., Kuramitsu H.K.;  
RT "Polymerase chain reaction for the detection of flaa-1 genes of oral  
RT Spirochetes in advanced periodontal pockets."  
RL Arch. Oral Biol. 0:0-0(2000).  
DR EMBL: AF132817; AAF64554.1; -  
FT NON-TER 1  
FT SEQUENCE 285 AA: 31917 MM: C64874F0387DBED8 CRC64:

Query Match 19.7%; Score 339; DB 2; Length 285;  
Best Local Similarity 30.9%; Pred. No. 5e-16;  
Matches 87; Conservative 49; Mismatches 104; Indels 42; Gaps 6;

OY 38 VLDFAELARDPSSSTRDLDTNVDVYSGAGSIKPEDVNDIGINNMSVLLTPSARLQAV 97  
Db 21 VLDYASLAD-----TSYTD-----EQKALMRSSLAQAQVEVNLNSASRPVA 62  
OY 98 VKNSVAPAVYKSESKRYAGDTILGVRVLPSPY-SOSSAMIMPFKIPFY----- 146  
Db 63 HAASRVLEAPVSEGAISFAGRGVLRVLEPTWDSNANNMIKPAFVIPAYEVMAQVDDQ 122  
OY 147 -----SGEGNOFLGKGLIDNKTKEIKYSVYSLGVEIDLEVLFDNMGMEYA 195  
Db 123 NVQAPPEEEKASGK-GRENGYGVYKNGVLAKSTAVNTYGMNYPHGLVMMRDQGEVHR 181  
OY 196 YSMGTLKFKGNADLIWSPNYIPNISRIKDDVFNPLASSKMFKAFRVSKSSKSVK 255  
Db 182 YFMGYLFDPSWKELVNPNPSYISDVRSREVR-LYPVYASPPHVVVEGFVWTRDAAHAG 240  
OY 256 NFIFYVKDLRVLYDKLSISDSDISESVFYKYEYSGTESLR 297  
Db 241 DVGIFKDKIITYDKAVLSTVRDEADDELWGIQARRRERKR 282

## RESULT 4

O9LA18 PRELIMINARY: PRT: 285 AA.

AC O9LA18: 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DR 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE flaa-1 (FRAGMENT).  
CN  
OS Treponema vincentii.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_TaxID=69710;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35580;  
RA Sato T., Kuramitsu H.K.;  
RT "Polymerase chain reaction for the detection of flaa-1 genes of oral  
RT Spirochetes in advanced periodontal pockets."  
RL Arch. Oral Biol. 0:0-0(2000).  
DR EMBL: AF132815; AAF64552.1; -  
FT NON-TER 1  
FT SEQUENCE 285 AA: 31830 MM: 4270CC4FAFA83C9F CRC64:

Query Match 19.6%; Score 337; DB 2; Length 285;  
Best Local Similarity 29.1%; Pred. No. 6,9e-16;  
Matches 89; Conservative 56; Mismatches 115; Indels 46; Gaps 8;

OY 13 ISTDVFAOETDGLAGSKRAEPGLVDFAEIARDPSSSTRDLDTNVDVYSGAGSIYKP 72  
Db 2 LNADIMADKSGGTMNRR-----TVLDYASLAD-----TSYTD-----EQK 37  
OY 73 EDMVVDLGINNMSVLLTPSARLQAVYKNSVAPAVYKSKSKRYAGDTILGVRVLPSPY-S 131

Db 38 ALMRSSLAVAQMEVLNNSARNPVAHAASRVLEAPVSEKAGSFAGERVLTGVRLFPTWDS 97  
QY 132 OSSAMIMPEKPIFEYS-----GSSGN-----QFL-GKGLDINIKMKRIKY 171  
Db 98 NAAAMKRPAYIAVEYMAVDQGNQAPTEEEKASGKGRFEDGGVGVNNGVLSTIAV 157  
QY 172 SVSLGVEIDLEVLFEEDMNGMEYASMGTLKFKGMADLINSNPYIPNISRIIKDVPN 231  
Db 158 NRYGMNYPHGLYVMBDQDEVRHRTFMGYLLFDSMKELVNNPSYISDVRSREVR-LYPV 216  
QY 232 YPLASSKMRKAFRVSSKSHSKVKNFIYVKDLRYLTKLSVSDSDISESVFKYETS 291  
Db 217 YPASTHYVVEGIMVTRDAHAAGDYGVYKFDVKIYDKAVLSTVRDFADEDLMGIDARR 276  
QY 292 GTESLR 297  
Db 277 EAERKR 282

RESULT 5  
006691 PRELIMINARY: PRT: 243 AA.  
AC 006691;  
DT 01-JUL-1997 (TREMBLrel. 04, Created)  
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)  
DE FLAA HOMOLOG-1.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_TaxID=160;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Porcella S.F., Radolf J.D., Norgard M.V.;  
RL Submitt. (Apr-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U97363; AAB63367.1; -  
SQ SEQUENCE 243 AA; 27299 MW; 89DA106A27100B41 CRC64;

Query Match 7.6%; Score 131.5; DB 2; Length 243;  
Best Local Similarity 22.5%; Pred. No. 0.11;  
Matches 56; Conservative 33; Mismatches 109; Indels 51; Gaps 8;

QY 48 PSTRLDITNYVDYV---YSGASGVKPEDMVVDLGINM---SVLLTPSARLQAYVN 100  
Db 21 PLAQSKSVNYQAYFFIDDFGAS-----EDQGL-AMRAAGSKRTTGFPLKTYFG 70  
QY 101 SVAPAVVSESKRYAGDTILGVRLFPYSOSSAMIMP-----PKIPYSGSGNQFL 155  
Db 71 MPQAVRMAGSMQCKDKKARFIVGECKFNROGNMMLDILPTKGSDEYIPLRGVVG- 126  
QY 156 GKGLDINIKMKRIKYSVSLGYEIDLEVLFEEDMNGMEYASMGTLKFKGMADLINSNP 215  
Db 127 -----FDVWVMGAGYQYSLEALVRCCTGVRHTLLGNLDFQGMKNLSVSVPT 173  
QY 216 YIPNISRIIKDVPNYPPLASSKMRKAFRVSSKSHSKVKNFIYVKDLRYLTKLSVSI 275  
Db 174 HIRQTS-----RYLSAQLHSFVGFRIRTSFSERVDYVFYFQDFAL--ANMHI 221  
QY 276 DSDIDSESV 284  
Db 222 DFTIDGHEL 230

RESULT 6  
P96127 PRELIMINARY: PRT: 242 AA.  
AC P96127;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE OUTER MEMBRANE PROTEIN.  
GN TROMP2.  
OS Treponema pallidum.

OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_TaxID=160;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NICHOLS;  
RX MEDLINE=97175551; PubMed=9023206;  
RA Champion C.I., Blanco D.R., Exner M.M., Erdjument-Bromage H.,  
RA Hancock R.E., Tempst P., Miller J.N., Lovett M.A.;  
RT "Sequence analysis and recombinant expression of a 28-kilodalton  
Treponema pallidum subsp. pallidum rate outer membrane protein  
(tomp2).";  
RL J. Bacteriol. 179:1230-1238(1997).  
DR EMBL: U65743; AAB47846.1; -  
SQ SEQUENCE 242 AA; 27234 MW; 86C71D3136862AF8 CRC64;

Query Match 7.3%; Score 125.5; DB 2; Length 242;  
Best Local Similarity 22.4%; Pred. No. 0.28;  
Matches 55; Conservative 31; Mismatches 104; Indels 55; Gaps 8;

QY 48 PSTRLDITNYVDYV---YSGASGVKPEDMVVDLGINM---SVLLTPSARLQAYVN 100  
Db 21 PLAQSKSVNYQAYFFIDDFGAS-----EDQGL-AMRAAGSKRTTGFPLKTYFG 70  
QY 101 SVAPAVVSESKRYAGDTILGVRLFPYSOSSAMIMP-----PKIPYSGSGNQFL 155  
Db 71 MPQAVRMAGSMQCKDKKARFIVGECKFNROGNMMLDILPTKGSDEYIPLRGVVG- 126  
QY 156 GKGLDINIKMKRIKYSVSLGYEIDLEVLFEEDMNGMEYASMGTLKFKGMADLINSNP 215  
Db 127 -----FDVWVMGAGYQYSLEALVRCCTGVRHTLLGNLDFQGMKNLSVSVPT 173  
QY 216 YIPNISRIIKDVPNYPPLASSKMRKAFRVSSKSHSKVKNFIYVKDLRYLTKLSVSI 275  
Db 174 HIRQTS-----RYLSAQLHSFVGFRIRTSFSERVDYVFYFQDFAL 217  
QY 276 DSDID 280  
Db 218 NMHI 222

RESULT 7  
083669 PRELIMINARY: PRT: 242 AA.  
AC 083669;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE OUTER MEMBRANE PROTEIN, PUTATIVE.  
GN TP0663.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_TaxID=160;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NICHOLS;  
RX MEDLINE=9832770; PubMed=9665876;  
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
RA Venter J.C.;  
RT "Complete genome sequence of Treponema pallidum, the syphilis  
spirochete.";  
RL Science 281:375-388(1998).  
DR EMBL: AE001240; AAC65636.1; -  
DR TIGR: TP0663; -  
KW Complete proteome.  
SQ SEQUENCE 242 AA; 27257 MW; 86C71D313FFECDF6 CRC64;

Query Match	7.1%	Score 121.5	DB 2	Length 242
Best Local Similarity	22.0%	Pred. No. 0.54		
Matches	54	Conservative	31	Mismatches 105; Indels 55; Gaps 8
48	PSSTLTDLTNNVDYV---	YSGASGIVKEPDMVVDGINNM---	SVLLTPSARLQAYKN	100
21	PLAQRKRVNVOAYRTIDPFGAS-----	EDGGL-AMRAAGSKFITKGPILKTFPG	70	
101	SVNAPAVVKSERYAGDTLIGRVLPFSYSSOASAMIP----	PKIPFYGESGNPL	155	
71	MPQAVRMAGSMQCKEAKRFISGECKFNRNOCNNMIDLIPTKGSGSYETLPRGVSG----		126	
156	GGKGLDNRKTKKEIKIVSYYSYGIEDLTVLPEDNMGMETAYSMGLTKFKGMADLIMSNPN	215		
127	-----FDVWGWAGYQYSLSLAYRDCYGRVHTLLIGLNDPFGMNLISVPT	173		
216	YIPNISRIIKDDVNNYPLASSKMFKAFRVSKSHSKKNEIFYVKDKIRLVYDKLSVSI	275		
174	HIPQVS-----RYLSGASQHLSPFVGRIRITXTSERVDY-----	VFDDQFKALA	217	
276	DSDID 280			
218	NMHID 222			
RESULT	8			
09ZHL3	PRELIMINARY;	PRT;	4152 AA.	
ID	09ZHL3			
AC	09ZHL3			
DT	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	LARGE SUPERNATANT PROTEIN 1.			
GN	LSPA1.			
OS	Haemophilus ducreyi.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;			
CC	Haemophilus.			
OX	NCBI_TaxID=730;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-35000;			
RX	MEDLINE=99030326; PubMed=9811662;			
RA	Ward C.K.; Lumley S.R.; Latimer J.L., Cope L.D., Hansen E.J.;			
RT	"Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.";			
RL	J. Bacteriol. 180:6013-6022(1998).			
DR	EMBL; AF057695; AAC79757.1; -			
DR	InterPro: IPR000130; Zn_MTPeptidse.			
DR	SMART: SM00235; ZmC; 1_			
SO	SEQUENCE	4152 AA; 456179 MW; 7D82DEDC988AB8F3 CRC64;		
Query Match	6.5%	Score 112;	DB 2;	Length 4152;
Best Local Similarity	22.8%	Pred. No. 1.6e+02;		
Matches	84	Conservative	44	Mismatches 133; Indels 108; Gaps 15
4	RAKSLIFLLSTVLEFAQETDGLAEGSKRAPELVLDFAELARDPSSTRLDTNNVYVY	63		
12	KVKNCILVPAENIKASGNSGSSNSKIAEDQ-----	EEDPSLACLS	55	
64	SCASGIVPEDMVVDLGINMSV-----	LTSPARKQA-----	Y	97
56	-----PLSSSHLGLNHSPLVYFKGSLSYLLSLMPTPLPLAQOYNAEALNGKY			107
98	VKNSVAVPVSSEKRRVAGDTILGVRVLPFSYSSOASAMIPFKIP-----	FYSG--ESG	151	
108	VDQSHTSRITVQCKTNDNSKDGIVVEIYANPEVDGSDNRKFEFNIPNSAVFNNSRTEST	167		
152	NOFLGKGLDNRKTKKEIKIVSYYSLGIEDLTVLPEDNMGMETAYSMGLTKFKG--	WADI	210	
168	SQLVAK-LHANIQLQKKA-----	LIINQVGDHESNIGALEVAGKADDI	213	
211	WSNPVYIPNISRIIKDDVNNYPLASSKMFKAFRVSKS----	HSSKVNFIIFYVKDLV	266	

```
Db      214 IYNPGITLNGVKRTINTD-----RFVSTSDILIPR---ENGLLSVRNKKV   256
Qy      267 LYDKLSVS-----IDSDISESVKRYEYSGTESLRLLKAHEFF-----KRVLKR   312
Db      257 TIDKGVAATNGLSHPEVVARNIDGKITVAKTENOKSVN--PANITEAAGSLMYNLKTR   314
Qy      313 EKISIAEGS 321
Db      315 EATPISSGT 323

RESULT          9
ID        Q9ZGS55 PRELIMITARY; PRT; 650 AA.
AC        Q9ZGS55;
DT       01-MAY-1999 (TREMblrel. 10, Created)
DI       01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT       01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE       HEAT SHOCK PROTEIN 70.
GN       HSP70.
OS       Arabidopsis thaliana (Mouse-ear cress).
OC       Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
OC       Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, Rosidae;
OC       eurosids II, Brassicales, Brassicaceae, Arabidopsids.
OX       NCBI_TaxID=3702;
RN       [1]
RP       SEQUENCE FROM N.A.
RC       TISSUE=GREEN SILIQUES;
RA       King K.;
RL       Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RR       [2]
RS       SEQUENCE FROM N.A.
SC       TISSUE=GREEN SILIQUES;
TA       Hindenhofer K., Praendl R., Schoeffl F.;
TT       "Seed-maturation-induced subset of heat shock protein mRNAs and heat-
RT       shock-element-binding protein complexes are dependent on ABI3 in
RL       Arabidopsis thaliana.";
RM       Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR       EMBL; AJ002551; CAA05547.1; -.
DR       HSSP; P08109; ICER.
DR       InterPro: IPRO01023; HSP70.
DR       Pfam; PF00012; HSP70_1.
DR       PRINTS; PR00301; HEATS_HOCHK70.
DR       PROSITE; PS00297; HSP70_1; 1.
DR       PROSITE; PS00329; HSP70_2; 1.
DR       PROSITE; PS01036; HSP70_3; 1.
KW       Heat shock.
SQ      SEQUENCE    650 AA; 71053 MW; 79A5B0C9CCCE13E CRC64;

Query Match               6.5%; Score 111.5; DB 10; Length 650;
Best Local Similarity     19.7%; Pred. No. 11;
Matches    69; Conservative 58; Mismatches 131; Indels 93; Gaps 14;
```



```
DB 280 LSTAGTTEIDSLFEGIDFYTTITRAREELNMDLFRKCMPEVEKCLDADKDKSSVMD 339
QY 287 VYETSGTESLRKLAHEFTFRVLAKEKISIAEGSFONFEKIESEKPEES 337
DB 340 VVVVGGSS-----TRIPKVOOLVO-----DEFNCKELCKSINPDEA 374

RESULT 10
Q9ZIQ2 PRELIMINARY; PRT; 1805 AA.
AC Q9ZIQ2;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE RTX PROTEIN (APXIV).
GN APXIVA OR APXIV.V1.
OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Actinobacillus.
OX NCBI_Taxid=715;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4074;
RA Kuhn R., Anderson T.J., Segers R.A.P.M., MacInnes J.I., Nicolet J.,
  Frey J.;
RT "Characterization of a new RTX determinant (apxiv) of Actinobacillus
  pleuropneumoniae."
RT Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=4074 (SEROTYPE 1 REFERENCE STRAIN);
RA Frey J., Segers R.P.;
RT "Live attenuated bacteria of the species Actinobacillus
  pleuropneumoniae."
RT Patent number EP0875574-A/5, 04-NOV-1998. AKZO NOBEL NV (NL).
DR EMBL; AF021919; AAD01698.1; -.
DR EMBL; AX002409; CAB77145.1; -.
DR InterPro: IPR001343; Hemlyns_Ca_bind.
DR InterPro: IPR003355; RTX_N.
DR Pfam: PF00353; hemolysinCbind; 5.
DR PRINTS; PR00313; CABNDNGRPT.
DR PROSITE; PS00330; HEWOLYSIN_CACTIUM; UNKNOWN_2.
SQ SEQUENCE 1805 AA; 202129 MW; A9076E3B351BD16 CRC64;

Query Match 6.4%; Score 110; DB 2; Length 1805;
Best Local Similarity 21.5%; Pred. No. 64;
Matches 68; Conservative 51; Mismatches 114; Indels 84; Gaps 17;

QY 22 TQGLAEGSRARPEGLVDFAE---LARDPSSTR---LDLTNYVDVYSGASIVAPEDM 75
DB 1008 TELNRNSTRKQSPENALYDLSELYSFNDPTKMEGLLSRLTID--YAKAOGFEY---- 1061
QY 76 VVDLGINNMSVLTTSARLQAYVKNSSVA---PAVYSESKRYAGDTLLGVLPSPYS 131
DB 1062 -----NM-----AATSNLTARLREAGVIFPESDIDKD----- 1090
QY 132 QSSAMIMPFKIPFYSGESGNOFL--GKGLIDINKTKMEIKVSVSLGYEIDLEVLPEDM 189
DB 1091 EKNNTLLGSKQNNNDLSGAGDILLGEGE--NDTLKSGYADYVIFSKHGQD--IYEDPT 1147
QY 190 NCMEXAVYSGTLLKFKGMALIMSNPIYINISRIIKDDVPNPPLASSKMRFAFVSKS 249
DB 1148 NMDNRARDIDITLKF-----TDVNY--AEVKKFRVNDL-----MLFGYHDT 1186
QY 250 HSSKVKNFIFYVLDRLVLYDKLSVSDIDSESVFK-----VYETSGTESLRKLAHEFT 305
DB 1187 DSVTVKSFYSVHD---YQDFKLEFA--DRSITRDELKAGLHLGYGTGND--IKDHADM 1239
QY 306 KRVLAKEKISIAEGSF 322
DB 1240 DSILEGKGNIDILRGY 1256
```

```
RESULT 11
Q9TTJ4 PRELIMINARY; PRT; 906 AA.
AC Q9TTJ4;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE INTER-ALPHA-TRYPSIN INHIBITOR HEAVY CHAIN H1.
GN ITI HCL.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Suzuki A.;
RT "The sequence of rabbit Inter-alpha-trypsin inhibitor heavy chain H1
  cDNA."
RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB035707; BAA88322.1; -.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR002035; WMFA.
DR Pfam: PF00092; vwa; 1.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; UNKNOWN_1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 906 AA; 100243 MW; 02D9D9DF585AB7BF2 CRC64;
```

```
Query Match 6.4%; Score 109.5; DB 6; Length 906;
Best Local Similarity 18.9%; Pred. No. 25;
Matches 78; Conservative 67; Mismatches 130; Indels 137; Gaps 19;

QY 1 MKRKASILFFLSTVLAQETDGLAEGSKRAPEGLVDFAELARDPSSTRDLTNYVD 60
DB 281 LKMSKSLFVI-----DISGMEGQVKQTKELLKL----- 314
QY 61 VYVSGASIVKPED---MYVDLGINNMSVLTSPS--ARLQA---YKNSVAPAVYKSE 111
DB 315 -----GDIRPEDYFDLYFGSRVOSWRSGLVPASEANLQARDYQRFSLGAT---- 363
QY 112 SKRYAGDTLLGVVL-----PFSYSSSAMIMPFKIPFYSGESGNOFLGKGLIDINKT 165
DB 364 --NLNGCLRLGRIETLLNNAQGNLPAVSKHAITL-----IMLTGEP-----TEGVTRPQI 411
QY 166 MKEIKVSV-----YSLGYEIDLEVLFEEDMNGMEYAVSMGTLKFGMADLIWSN----- 213
DB 412 LKNIRSAIGRFPYLSLGFCHDIDFNLKLSMEN-----NGMAQRITIEDHDAQO 462
QY 214 -PNYITNISRIIKDDVPNP---LASSKMRKAF-----RVSKSHSKYKNFT 258
DB 463 LOGFYQVANPLLYVDVDELLYPQDAVVALTQHRIKQYDGSSEIYVAGRIADHKIGSPK--- 519
QY 259 FYYKDLRLVLYDKLSVSDIDSESVFKYVETSG---TESLRKLAHEFTKRVL----- 309
DB 520 ---ADVRRAGEGOEFQTTCLVDEEEMKKLLREXGMLHNHVERLMAYLTIIQELARHYVA 576
QY 310 KLEKISIAEGSFQ-----NFVEKIES-----EKPEESP 339
DB 577 KGEKANVSSEALKMSLAVQFVPLTSMTRGMADEGLEPTIDKRPEDSQP 628

RESULT 12
Q9LHAB PRELIMINARY; PRT; 650 AA.
AC Q9LHAB;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 70 KDA HEAT SHOCK PROTEIN (HEAT SHOCK PROTEIN 70).
GN T2E22.11.
```

OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
NC NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
RT Sequence features of the regions of 4,251,695 bp covered by ninety pl.  
RT TAC and BAC clones.";  
RL DNA Res. 7:217-221(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV, COLUMBIA;  
RA MEDLINE=21016720; PubMed=11130713;  
RX Salanoubat M., Lemcke K., Rieger M., Ansoirge W., Unselde M.,  
FA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermayer B.,  
RA Delesny M., Boutry M., Grivell L.A., Macho R., Pulgomech P.,  
RA De Simone V., Choise N., Artiguenave F., Robert C., Brottier P.,  
RA Wincker P., Catolico L., Weissenbach J., Saurin W., Queller F.,  
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Bernes V.,  
RA Wunmbach E., Drzonex H., Erile H., Jordan N., Bangert S.,  
RA Wedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,  
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,  
RA Conrad A., Hornschfer K., Kauer G., Loehner T.-H., Nordiek G.,  
RA Reichelt J., Schärfer M., Schoen O., Barges M., Terol J., Climent J.,  
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,  
RA Cooke R., Lande M., Berger-Liauro C., Purrelle B., Masny D.,  
RA De Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,  
RA Monfort A., Argilheu A., Flores M., Liguori R., Vitale D.,  
RA Mannhaupt G., Haese D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,  
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
RA Rooney T., Rizzo M., Walts A., Uteback T., Fujil C.Y., Shea T.P.,  
RA Creasy T.H., Haas B., Malt R., Wu D., Peterson J., Van Aken S.,  
RA Pal G., Mlltscher J., Sellers P., Gill J.E., Feldblum T.V.,  
RA Preuss D., Lin X., Nienan W.C., Salzberg S.L., White O., Venter J.C.,  
RA Fraser C.M., Kaneo T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,  
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakayama S., Nakazaki N., Shimo S., Takeuchi C., Wada T.,  
RA Nakabe A., Yamada M., Yasuda M., Tabata S.;  
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
thaliana";  
RL Nature 408:820-822(2000).  
DR EMBL; AP002055; BAB02269.1;  
DR EMBL; AC069474; AAG51030.1;  
DR InterPro; IPR001023; HSP70.  
DR Pfam; PF00012; HSP70; 1.  
DR PRINTS; PR00301; HEATSHOCK70.  
DR PROSITE; PS00297; HSP70.1; 1.  
DR PROSITE; PS00329; HSP70.2; 1.  
DR PROSITE; PS01036; HSP70\_3; 1.  
KW Heat shock.  
SQ SEQUENCE 650 AA; 71101 MW; 33381E96461C5B9 CRC64;

Query Match 6.3%; Score 108.5; DB 10; Length 650;  
Best Local Similarity 19.7%; Pred. No. 18;  
Matches 69; Conservative 58; Mismatches 131; Indels 93; Gaps 14;

OY 43 ELARPSSRLDITNYYVYVYGAGIYKPEEDMVDLGINNW-SVLTLPASRLQAYYKN 100  
DB 61 OYANPNTIVDAKRLIGRIS-----DPSVQADKSHMPFVYVSGPK----- 104  
OY 101 SVVAPAVV---KSESKRYAGDTILIGRVLPSPYSOSSAMIMPPK-----IPYSGESGN 152

DB 105 ----PMTVNHKEGKQSAEISSM-VLTMRLEAFASLPKNAVYIPAYFANDSR 159  
OY 153 QFL-GKGLIDINKMKET-KVSYSLGEYDLE-----VLPEDNMGMEYASMKOT 202  
DB 160 QATDAGVYSGLVNMRINEPTAAIAVGLDKKASVGEKNVLLFDGGGFDVSLTLIE 219  
OY 203 -----FKGADLIWSPNPIPIINSSRIIDDPVNYPLASSKMFARFYSKS 249  
DB 220 EGIEFVKATAGDTLHGDEFDNRVNHVQEFKRRKKKIDYGNRALRLRTACERAKRT 279  
OY 250 HSKVKNFI-----FVYKDLRVLYDKLSVI-----PSDIDESVFK 286  
DB 280 LSTQAQTTIEIDSLFEGIDFTTTTRAFEEELNDLFKCKMEPEVKCIRDAKMKSSVHD 339  
OY 287 VYENSGTESLRKLKAHETFKRVALKREKISIAEGSFQNFVERISEKPEES 337  
DB 340 VVLVIGGS-----TRIPVQQLQ---DFNGKRLGCSINDEA 374

RESULT 13  
OYHK03  
ID ID PRELIMINARY; PRT; 1124 AA.  
AC OYHK03;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE HYPOTHETICAL MEMBRANE PROTEIN.  
GN TA0543.  
OS Thermoplasma acidophilum.  
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;  
OC Thermoplasma.  
NC NCBI\_TaxID=2303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 1728;  
RX MEDLINE=20479972; PubMed=11029001;  
RA Rupp A., Grail W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;  
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma  
acidophilum";  
RL Nature 407:508-513(2000).  
DR EMBL; AL445064; CAC11683.1;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1124 AA; 126534 MW; ADCEC272CC3FC2C9 CRC64;

Query Match 6.3%; Score 107.5; DB 1; Length 1124;  
Best Local Similarity 22.0%; Pred. No. 48;  
Matches 76; Conservative 38; Mismatches 95; Indels 137; Gaps 18;

OY 25 LAEGSKRAEPGE-----LVLPFAELARPSSRLDITN---YVDY-----VY 63  
DB 809 LINGSTYVPSGINSISLTSGIYNITAINLSTYLRNFRITDYLSTTLIDYIRYSKVY 868  
OY 64 SGASGIVKPEEDMVDLGINNWSVLTLPASRLQAYYKNSVAPVYKSESKRY---AGDTI 120  
DB 869 INASGYADEFE-----IKNNSYVYVAGDTI 895  
OY 121 LGVRLFPYSOSSAMIMPPKIFP--YSGESGN-----QFLGKGL-IDNINKTMEIKVS 172  
DB 896 -----HIPEGLYSIRAEVQGYSEYGNKISIDSSAIIINTFS 934  
OY 173 ----VYSLGEYDLEVLPEDNMGMEYAYSMGTLKFKGADLIWSPNPIPIINSSRIIRD 228  
DB 935 RLRFMHTATDIPADIVG--NTRFYGONFSLVDGYNDLSISKRG----- 980  
OY 229 VPNYPLASSKMFARFYSKSSKVKNFIFYVKDLR-VLYDKLSVSDSDISESEV-- 285  
DB 981 ---YPISESSV-----LNSNMTLEFA-----LRAILV---KITLEADVDSFTIYAG 1020  
OY 286 -KVYETSGTESLRKLKAHETFRVRVLKAEKISIAEGSQNFVEKE 330

Db 1021 NKTYSVSGNDLTIYLPYGE-----HSIIQKDGYSNIFETID 1057

RESULT 14

ID 014789 PRELIMINARY: PRT: 947 AA.

AC 014789:

DT 01-JAN-1998 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE BRDT.

GN Homo sapiens (Human).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98035876; PubMed=9367677;

RA Jones M.H., Numata M., Shimane M.;

RT "Identification and characterization of BRDT: A testis-specific gene related to the bromodomain genes RING3 and Drosophila fsh.";

RL Genomics 45:529-534(1997).

DR EMBL; AF019085; AAB87862.1; -.

DR InterPro: IPR001487; Bromodomain.

DR Pfam: PF00439; bromodomain.2.

DR PRINTS: PR00503; BROMODOMAIN.

DR PROSITE: PS00633; BROMODOMAIN\_1; 2.

DR SMART; SMO0297; BROMO; 2.

DR SMART; SMO0297; BROMO; 2.

SO SEQUENCE 947 AA; 107891 MW; EC6AE07C45863783 CRC64;

Query Match 6.2%; Score 107; DB 4; Length 947;

Best Local Similarity 18.5%; Pred. No. 40;

Matches 76; Conservative 80; Mismatches 120; Indels 134; Gaps 20;

QY 1 MKRKKSLIFLLSTVLFQNDGLAEGSKRAPEGLVDFAELADPSST-RDLDTNIV 59

Db 118 MDALEKLEPMOKLSQMPQEOYGVAKERIKGTQONIAVSAKESPSATEKVFQOET 177

QY 60 DYVYSGASGIVKPEDMVVDLGINMSVLLTPSARLQAYVK---NSVPAVAVKSESKRY 115

Db 178 PSVFPRTS--ISPLANYQASVNSSS---QTAQVTKGVRKADTTTPATSAVKASSE-- 230

QY 116 AGDTILGVVLPSPYSQSSAMTMPK-----IPFYGESGNOFLGGLDINIKTME 168

Db 231 -----FSPTFEKES-VALPIKENMPKVLTP---DSQOQ---NVVETVAVTEQ 272

QY 169 IK-----VSYSLG---YEL-----DLEVLFEDEMNGMEY- 194

Db 273 LRHCSEILKEMAKKHFSYAMPFPYNPDVNAIGLANYVKNPKMLDGLTIKEMDQEKY 332

QY 195 -AVSMGTLFKGMADLIMSN-----PNYIPNISRIIKDDVNPYPLASSKMRKAFRVS 247

Db 333 DAYS-----FAADVRLMFNCKYKNPDPHEVYTMARMLD-----VF 369

QY 248 KSHSSKVKNFIFYKDLRLVYDKLSVSDSIDSESVFVYETSGTESLAKIKAH----- 302

Db 370 ETHFSKIP--IEPVESMPLCIYIKTDIT-----ETTGRENTNEASSEGNSSD 413

QY 303 -----EFFKVLKLEKIS-----IAGSFONFEVKEIESEKPESSPK 340

Db 414 DSEDERVKRLAKLQEQLKAVHQOQLOVLSQVPPFKLNKKKSKKKKKKK 463

RESULT 15

ID 09KN45 PRELIMINARY: PRT: 1181 AA.

AC 09KN45:

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

DE ICMF-RELATED PROTEIN.

GN VCA0120.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI\_Taxid=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EL TOR N16961 / SEROTYPE O1;

RX MEDLINE=20406833; PubMed=10952301;

RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,

RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,

RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,

RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,

RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,

RA Fraser C.M.;

RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio

RT cholerae.";

RL Nature 406:477-483(2000).

DR EMBL; AE004353; AAF96034.1; -.

DR TIGR; VCA0120; -.

KW Complete proteome.

SO SEQUENCE 1181 AA; 134777 MW; 3EE262B2B0FE7949 CRC64;

Query Match 6.2%; Score 106; DB 2; Length 1181;

Best Local Similarity 23.3%; Pred. No. 65;

Matches 77; Conservative 54; Mismatches 118; Indels 82; Gaps 17;

QY 28 GSKRAPEGLVDFAEL-----ARPPSTRDLDTNIVVDVYS---GASG 68

Db 866 GSKPAIYTEVLASVDLKSITKSIQDAPDVGMAALDITAKRVLVN-ADPIYTLKRSSG 924

QY 69 IYKPED-MVVDLGINMSVLLTPSARLQAYVKNVY--APAVVSKSESKRYAG----- 117

Db 925 LRPFLDSMAKLADESMYVY-----KQEAIRKLEVRWTEDEVYKTFQSKLAGRYFNPAS 978

QY 118 --DTILGVRLF--PS-----YSQSSAMTMPKFPFYSGESGNOFLGGLDINIKTME 167

Db 979 NKDVALADEFAFPAPNGTYDINFNOQKMFIDE-NISVASDSDSAOSITIRREVLDQIKQAO 1037

QY 168 EIVSYVSLGCEIDLELFE-----DMNGMEYAVNGTLKFGMADLIMSNP 214

Db 1038 KIRAEAFNKRGLIDVFSVSEPLSLNNKRSYLVNDQGLAYSHGP---RENNELLTPNT 1094

QY 215 -----NYTP--NISRIIKDDVNPYPLASSKMRKAFRVS KSHSSKVKNFIFYYK 262

Db 1095 LRDSAVSKVTLIPTQTNMSPRSDQ---IQGPMAFFRLDDGDIVVASAOTS--VDFKFIYD 1149

QY 263 DLRLVLYDKLSVSDSIDSESVKVYETSGT 293

Db 1150 GGEMTY-RINAEADAPFTEFLKSEKLSKT 1179

Search completed: March 20, 2002, 00:00:37

Job time: 2423 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 19, 2002, 22:15:13 ; Search time 63.46 Seconds  
(without alignments)  
120.921 Million cell updates/sec

Title: US-09-004-395-2  
Perfect score: 1719  
Sequence: 1 MKRRAKSLFELSTVLFQAQ.....FQNFVEKIESKPESSPKN 341

Scoring table:  
BIOSDM62  
Gapol 10.0 , Gapept 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PTC05.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	125.5	7.3	242	1	US-08-599-480-4
2	110	6.4	1657	3	US-09-057-570-2
3	110	6.4	1805	3	US-09-057-570-7
4	97.5	5.7	530	4	US-08-975-762-73
5	97.5	5.7	530	4	US-09-295-028-73
6	97.5	5.7	530	4	US-09-106-582-73
7	97.5	5.7	546	4	US-09-066-047-2
8	97.5	5.7	590	4	US-08-975-762-64
9	97.5	5.7	590	4	US-09-295-028-64
10	97.5	5.7	590	4	US-09-106-582-64
11	94	5.5	1382	3	US-09-057-570-4
12	93	5.4	607	3	US-08-486-099-116
13	93	5.4	607	3	US-08-484-223B-116
14	93	5.4	607	3	US-08-475-668A-116
15	93	5.4	607	3	US-08-475-668A-116
16	93	5.4	607	3	US-08-485-551A-116
17	93	5.4	607	3	US-08-471-913A-116
18	93	5.4	607	3	US-08-485-264A-116
19	93	5.4	608	3	US-08-360-107A-126
20	92	5.4	679	1	US-08-441-139-5
21	91.5	5.3	654	4	US-09-090-912-2
22	91	5.3	1861	2	US-08-790-912-4
23	90	5.2	978	2	US-08-415-593-43
24	90	5.2	1186	1	US-08-485-568A-4
25	90	5.2	1186	1	US-08-357-698-6
26	90	5.2	1186	2	US-08-590-554A-4
27	90	5.2	1186	2	US-09-184-223-4

28	90	5.2	1186	5	PCT-US93-12682-6	Sequence 6, Appl1
29	88.5	5.1	1180	4	US-08-660-148-2	Sequence 2, Appl1
30	88.5	5.1	1212	4	US-08-660-148-5	Sequence 5, Appl1
31	87.5	5.1	627	1	US-08-703-947-2	Sequence 2, Appl1
32	87.5	5.1	877	1	US-08-072-574-12	Sequence 12, Appl1
33	87.5	5.1	877	3	US-08-486-270-12	Sequence 12, Appl1
34	87.5	5.1	877	3	US-08-367-264-12	Sequence 12, Appl1
35	87.5	5.1	1180	1	US-08-072-574-8	Sequence 8, Appl1
36	87.5	5.1	1180	1	US-08-486-270-8	Sequence 8, Appl1
37	87.5	5.1	1180	3	US-08-367-264-8	Sequence 8, Appl1
38	87.5	5.1	1212	1	US-08-072-574-10	Sequence 10, Appl1
39	87.5	5.1	1212	1	US-08-486-270-10	Sequence 10, Appl1
40	87.5	5.1	1212	3	US-08-367-264-10	Sequence 10, Appl1
41	87	5.1	678	1	US-08-844-085-2	Sequence 2, Appl1
42	87	5.1	715	2	US-08-484-993B-10	Sequence 10, Appl1
43	87	5.1	715	2	US-08-484-158B-10	Sequence 10, Appl1
44	87	5.1	715	2	US-08-484-596A-10	Sequence 10, Appl1
45	87	5.1	715	2	US-08-480-150A-10	Sequence 10, Appl1

ALIGNMENTS

RESULT 1  
US-08-599-480-4  
; Sequence 4, Application US/08599480  
; Patent No. 5735459  
; GENERAL INFORMATION:  
; APPLICANT: Blanco, David R.  
; APPLICANT: Miller, James N.  
; APPLICANT: Lovett, Michael A.  
; APPLICANT: Champion, Cheryl I.  
; APPLICANT: Tempst, Paul J.  
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF A  
; TITLE OF INVENTION: T. Pallidum RARE OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/599,480  
; FILING DATE: 23-JAN-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Learn, June M.  
; REGISTRATION NUMBER: 31,238  
; REFERENCE/DOCKET NUMBER: 07419/018001 (CIP of 016001)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 678-5070  
; TELEFAX: (619) 678-5099  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 242 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-599-480-4  
Query Match 7.3%; Score 125.5; DB 1; Length 242;  
Best Local Similarity 22.4%; Pred. No. 5.1e-05;  
Matches 55; Conservative 31; Mismatches 104; Indels 55; Gaps 8;  
QY 48 PSSRIIDLTVYVYV--YSGASGVKPEDMYVDLGNM-----SVLLTFSARLQAVKN 100

Db 21 PAAORSVKNTQAFIDDEFGAS-----EDOGL-AMRAAGSKFTIKGFPILKYEFG 70  
QY 101 SVAFAVAVKSESKRYAGDTLLGVRVLPFSYSSOSAMIMP-----PKPIPFYSGSGNQFL 155  
Db 71 MPOAVRMAGSWGOKKREKRFVIGVECKEENRQNNMIDLPTKGSDEYELPLRGVSG---- 126  
QY 156 GKGLDINKTKMEIKVSVYSGLYEIDLEVLFPEDMNGMEYASMGTLKFKGMADLIWSNPN 215  
Db 127 -----PDVWVMGAGYQYSLEALVYDCRGVHTLLIGNLDPFQGMKNSVSVPT 173  
QY 216 YIPNISHTIKDDVPNYPLASSKMRKFAFRVSKSHSKYKNIFFVKDLRVLYDKLSVI 275  
Db 174 HLPQTS-----RYLGSQHLSEFVGFRJRTSPSERVDFY-----VFDDQFKALA 217  
QY 276 DSDID 280  
Db 218 NMHID 222

## RESULT 2

US-09-057-570-2  
; Sequence 2, Application US/09057570  
; Patent No. 6013266  
; GENERAL INFORMATION:  
; APPLICANT: Segers, Ruud P.A.M.  
; APPLICANT: Frey, Joachim  
; TITLE OF INVENTION: Live attenuated Actinobacillus  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Akzo No. 6013266el Patent Department  
; STREET: 1300 Piccard Drive, Suite 206  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/057,570  
; FILING DATE: 09-APR-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gormley, Mary E.  
; REGISTRATION NUMBER: 34,409  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 948-7400  
; TELEFAX: (301) 948-9751  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1657 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-057-570-2

Query Match 6.4%; Score 110; DB 3; Length 1657;  
Best Local Similarity 21.5%; Pred. No. 0.048;  
Matches 68; Conservative 51; Mismatches 114; Indels 84; Gaps 17;

QY 22 TDGLAEGSKRAPEGELVDFAE---LARDPSSTR---LDLTNYVDYVYSGAGIVKPEDM 75  
Db 860 TEALFNSTFKOSPENALDLESEYISFPNDPTMKRGILLISRYID--YAKAGGYE----- 913  
QY 76 VVDIGINNMVLLTPSARLOAVKNSVYA---PAVVSSESKRYAGDTLLGVRVLPFSYS 131  
Db 914 -----NW-----AATSNLTJARLREAGVIFAESTDLKGD----- 942  
QY 132 QSSAMIMPFKIPFYSGESGNQFL--GKGLDINKTKMEIKVSVYSLGEYIDLEVLPEDM 189

Db 943 EKNNTILGSKDNNSLGSAGDILLIGGEG--NDTLKGSYGADYIFRSKHGD--IYVEDT 999  
QY 190 NQMEVAYSNGTLKFGMADLIWSNPNYIPNISRIIKDDVPNYPLASSMRKARFVSKS 249  
Db 1000 NDNRRARDIDTLK-----TDVNY-AEVEFRVNDL-----MLGYYHDT 1038  
QY 250 HSSKYNFIFFVKDRLVYDKLSVSDSDIDSESVFK---VYETSGTESLRKLAHETP 305  
Db 1039 DSVTVKSFSSHND---YQDKELEFA--DRSITRDELKKGHLHYGTGDND---INDHADW 1091  
QY 306 KRYVLKREKISTAEGSF 322  
Db 1092 DSILEGKGNDILRGY 1108

## RESULT 3

US-09-057-570-7  
; Sequence 7, Application US/09057570  
; Patent No. 6013266  
; GENERAL INFORMATION:  
; APPLICANT: Segers, Ruud P.A.M.  
; APPLICANT: Frey, Joachim  
; TITLE OF INVENTION: Live attenuated Actinobacillus  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Akzo No. 6013266el Patent Department  
; STREET: 1300 Piccard Drive, Suite 206  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/057,570  
; FILING DATE: 09-APR-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gormley, Mary E.  
; REGISTRATION NUMBER: 34,409  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 948-7400  
; TELEFAX: (301) 948-9751  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1805 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-057-570-7

Query Match 6.4%; Score 110; DB 3; Length 1805;  
Best Local Similarity 21.5%; Pred. No. 0.057;  
Matches 68; Conservative 51; Mismatches 114; Indels 84; Gaps 17;

QY 22 TDGLAEGSKRAPEGELVDFAE---LARDPSSTR---LDLTNYVDYVYSGAGIVKPEDM 75  
Db 1008 TEALFNSTFKOSPENALDLESEYISFPNDPTMKRGILLISRYID--YAKAGGYE----- 1061  
QY 76 VVDIGINNMVLLTPSARLOAVKNSVYA---PAVVSSESKRYAGDTLLGVRVLPFSYS 131  
Db 1062 -----NW-----AATSNLTJARLREAGVIFAESTDLKGD----- 1090  
QY 132 QSSAMIMPFKIPFYSGESGNQFL--GKGLDINKTKMEIKVSVYSLGEYIDLEVLPEDM 189  
Db 1091 EKNNTILGSKDNNSLGSAGDILLIGGEG--NDTLKGSYGADYIFRSKHGD--IYVEDT 1147  
QY 190 NQMEVAYSNGTLKFGMADLIWSNPNYIPNISRIIKDDVPNYPLASSMRKARFVSKS 249

```

Db      1148 NNDNDRARDIOFLKFE-----TDVNI-AEVAFRRYDNL-----MLFCYHDT 1186
Qy      230 HSKAKKNITFYKDI.RVLTKLSVIDSDIDSESVPK---YYETSGTESLKKAKHETF 305
Db      1187 DSVTVRKSPYSVSHD---YOQDKLEFA-DRSITRDELIKAGLHLGTGDDGND---IKDHADM 1239
Qy      306 KVKLRKREKSIAEGSF 322
Db      1240 DSILEGGKNDILRGY 1256

RESULT          4
US-08-975-762-73
: Sequence 73, Application US/08975762
: Patent No. 6207169
: GENERAL INFORMATION:
:   APPLICANT: Reed, Steven G.
:   APPLICANT: Lodes, Michael J.
:   APPLICANT: Houghton, Raymond
:   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
:     NUMBER OF SEQUENCES: 73
: CORRESPONDENCE ADDRESS:
:   ADDRESSEE: SEED and BERRY LLP
:   STREET: 6300 Columbia Center, 701 Fifth Avenue
:   CITY: Seattle
:   STATE: Washington
:   COUNTRY: USA
:   ZIP: 98104
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/975,762
:   FILING DATE: 21-MAR-1997
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
:   NAME: Makl, David J.
:   REGISTRATION NUMBER: 31,392
:   REFERENCE/DOCKET NUMBER: 210121.439
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 206-622-4900
:   TELEFAX: 206-682-6031
: INFORMATION FOR SEQ ID NO: 73:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 530 amino acids
:     TYPE: amino acid
:     STRANDEDNESS:
:       TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-975-762-73

Query Match           5.7%; Score 97.5; DB 4; Length 530;
Best Local Similarity 22.8%; Pred. No. 0.15;
Matches 68; Conservative 54; Mismatches 101; Indels 75; Gaps 16;

Qy      74 DMVV-----DLGINMWSVLLTPSARLAQAYVKN-----SYVAPAVKSE 111
Db      110 DMVEKEFAHDIGIR-----VGSNSLRSLIKIRIFQDANGVFDEGRYEAVALDSGMTES 163
Qy      112 S-----KRVAADTTLGVAVLPFSYSOSSAMIMPPKPIPYSSGSGNQPLGKLIDNIKTM 166
Db      164 SYVNKIRNALPSTII-MECLEFPNRAE-----LHIPYDALADVYLGL-LQHRVADI 213
Qy      167 KEIKVSYSTL-GYEI---DLEVLFEDMNGMEYAYSMTGLFKFG-----WADLTW 211
Db      214 VEISSDAVDIGSGDISDELQKLFEE-----GYKNSLNPFEPYSAVIYIMAEDDLADIVTY 269
Qy      212 SNPNTIPISSRIKD--DVNPYPLA---SKMKRFKAPEVSKSSHSSKYANKETIFYVKD-- 264

```

[illegible]

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MakI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-106-582-73

```

Query Match	5.78;	Score	97.5;	DB	4;	Length	530;
Best Local Similarity	22.08;	Pred. No.	0.15;				
Matches	68;	Conservative	54;	Mismatches	101;	Indels	75;
						Gaps	16

```

OY 74 DMVY-----DLGINNWSALLTPSRLOAYVN-----SVWPAAYVXSE 111
      |||||      |||||      |||||      |||||      |||||
Db 110 DMVVEKFAHDGIR-----VGSLSLSLKINIFIDPDANGVFEQOEKYEAVLADSGITES 163
      |||||      |||||      |||||      |||||      |||||
OY 112 S-----KRYAGDTLLGVAVLPSPYSQSSAMIMPEFKIPYVSGESNOFLKGLIDNIKTM 166
      |||||      |||||      |||||      |||||      |||||
Db 164 SYVKKIRALPSTLL-MECLFPPNPAE-----LHLPYDALAKVDLGL-LQHRVADI 213
      |||||      |||||      |||||      |||||      |||||
OY 167 KEIKVSYVSL-GYEI--DLEVLREDNMGMEYASMGCLKFKG-----MAQLIW 211
      |||||      |||||      |||||      |||||      |||||
Db 214 VEISDAVIDIGSDISDEDLAKLEE---OYKSLNFPETRSADYITIMADELLADYIV 269
      |||||      |||||      |||||      |||||      |||||
OY 212 SNNPYIPNISRRIKID-DVNPYPLA---SSKMRKFAFRVSSHSKYNKFIYVKD- 264
      |||||      |||||      |||||      |||||      |||||
Db 270 SDOEVDVEIKNSELHODRDVNLFTFOKNEALYKAYQEGKSEFEELVSDAGYITIEDIAL 329
      |||||      |||||      |||||      |||||      |||||
OY 265 -RVLYDKLSVSDSDIDISESEFKYE---TSQESLRSKKAHTEKRRVL-KIUREKS 316
      |||||      |||||      |||||      |||||      |||||
Db 330 NNISKDVLPGVVRNVVPLNKGVESEMRPSYVGHMILVRKHHTIETEDLEKLEKKS 387
      |||||      |||||      |||||      |||||      |||||

```

RESULT 7  
 US-09-066-047-2  
 : Sequence 2, Application US/09066047A  
 : Patent No. 6306394  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: MURPHY, Cheryl  
 : STOREY, James  
 : BELTZ, Gerald A.  
 : COUGHLIN, Richard T.  
 :  
 : TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF  
 : USE OF GRANULOCYTIC EHRlichIA  
 :  
 : NUMBER OF SEQUENCES: 41  
 :  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: HALE AND DORR LLP  
 : STREET: 60 State Street  
 : CITY: Boston  
 : STATE: Massachusetts  
 : COUNTRY: United States  
 : ZIP: 02109  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,047A
FILING DATE: 24-APR-1998
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/044,869
FILING DATE: 25-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Superko, Colleen
REGISTRATION NUMBER: 39,850
REFERENCE/DOCKET NUMBER: 106,941,156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 526-6000
TELEFAX: (617) 526-5000
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 546 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-066-047-2

```

Query Match	5.7%	Score	97.5	DB	4	Length	546
Best Local Similarity	22.8%	Pred. No.	0	16			
Matches	68	Conservative	54	Mismatches	101	Indels	75
						Gaps	16

```

OY 74 DMVY-----DLGINNMSVLLTPSARLAAVYN-----SVAAVAVXSE 117
    ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db 11 DMVEKFAHDDGIR-----VGSNLSRLKLNINIFODANGVFOEGEREAVLADSGMTES 64
    ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
OY 112 S-----KRYAGDTLLGVRLFPYSOSAMIMPEPKLTFESGESNOFLKGLIDNIKTM 166
    ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db 65 SYVKNIRNALSTLL-MCELEPRNRAE-----LHJPYVALAKKADVLGL-LOHRVADI 114
    ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
OY 167 KEIKVSVSL-GYEL---DLFVLEFEDNMNGMEYAVSMGLKPKG-----WADLIW 211
    ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db 115 VEISSDAVDIGSDISDEDELOKLEEE-----OYKNSLNFPEKRSADYIIMAEDDLADVIY 170
    ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
OY 212 SNPNTIPNISRIIKD-DVBNPYPLA---SSKMRFKAFRYSKSHSKVKNFIFYVKD-- 266
    ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db 171 SDQEVDAVEIKSELHDQNDVNLVFTDKNENAEALVYKVAQESKSEELVSDAGYIETIDAL 230
    ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
OY 265 -RVLYDKLVSIDSDIDSESFFKYVE-----TSGRESLKLKLAKHETFRVY- KLEKXS 316
    ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db 231 NNISXDVLFPVGVVRNVVFLANGVEVSEMERSVYVGMHIMVILKKBHTEDEDEKLEKXS 288
    ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||

```

RESULT 8  
US-08-975-762-64  
Sequence 64, Application US/0895762  
Patent No. 6207169  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
City: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

**TREATM**



Query Match	5.7%;	Score 97.5;	DB 4;	Length 590;
Best Local Similarity	22.8%;	Pred. No. 0.18;		

```

110 DMVEKFAHDLGIR-----VGSNSLSRLIKNIRIFODANGVPEQDERYEAVLADSGTSES 163

```

OY 112 S-----KRYAGDTILGVRVLPFSYSSOSAMIMPFKIPFYSGSGNOFLGKGLIDIKTM 166  
DB 164 SYVNKRIRNALPSTIL-MELLFPNRFA-----LHIPTDALADVYIGL-LQHVADI 213  
OY 167 KEIKSVYSVL-GYEI--DLEVFEDMNGMEYAYSMGTLKFKG-----WADLIW 211  
DB 214 VEISSDAVIDSGDISDDELQKLFEE---QYKNSILNPEYRSADYIIMAEDDLADVIY 269  
OY 212 SNPNITPNISSRIKD--DVPNYPLA---SSKMRFAFVSKSHSKVNFIFYVLDL--- 264  
DB 270 SQOEVDVEIKNSLHDQKRVNLVFTDKNEADLAYKAYGEGSFEEVLSDAGTTELDIAL 329  
OY 265 -RVLVDKLSVDSIDSDISESVFKYVE---TSGTESIRKLKAHEFKRVL-KLREKIS 316  
DB 330 NMSKDVLPVGRVNYVFALNEGESEMFPSYGVGMHMKYIRKHETKEDLEKIKKIS 387

RESULT 11  
US-09-057-570-4  
Sequence 4, Application US/09057570  
Patent No. 6013266  
GENERAL INFORMATION:  
APPLICANT: Segers, Ruud P.A.M.  
APPLICANT: Frey, Joachim  
TITLE OF INVENTION: Live attenuated Actinobacillus  
TITLE OF INVENTION: pleuropneumoniae  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Akzo No. 6013266el Patent Department  
STREET: 1300 Piccard Drive, Suite 206  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/057,570  
FILING DATE: 09-APR-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Gormley, Mary E.  
REGISTRATION NUMBER: 34,409  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 948-7400  
TELEFAX: (301) 948-9751  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1382 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-057-570-4

Query Match 5.5%; Score 94; DB 3; Length 1382;  
Best Local Similarity 20.5%; Pred. No. 1.7; Indels 84; Gaps 17;  
Matches 65; Conservative 52; Mismatches 116;

OY 22 TDGLAEGSKRAPEGLVLDFAE---LARDPSSTR--LDLTNVVDVYSGASGIVPEDM 75  
DB 860 TEALFNSTYKOSPENVLYLDRELYSTFNDPTMKESGLILSRID--YAKAGCFYE----- 913  
OY 76 VVDLGINNSVLLTPSARLOAVYKNSVA---PAVYKSESKRYAGDTILGVRVLPFSYS 131  
DB 914 -----NM-----AATSNLTIRAREAGVICAESTDLKGD----- 942  
OY 132 QSSAMIMPFKIPFYSGSGNOFL--GKGLIDIKTMKIKSVYSLGIEIDLEVLFEEDM 189  
DB 943 EKNNTVLGSGOKNNLSGSAGDDLIGGEG--NDTLKSGYADTYIFSGHGOD--VITYEYS 999

OY 190 NMEYAYSMGTLKFKGMADLIMSNPNITPNISSRIIKDDVPNYPLASSKMRFAFRVSKS 249  
DB 1000 DSANSKKDIDTLKF-----TDVNY-AEVKFRARDNL-----MELGYHDT 1038  
OY 250 HSSKVKNFIFYVVDLRLVDKLSVDSIDSDISESVFK---VYEGSGTESLKLKAHEFE 305  
DB 1039 DSVTVKSPFSYSHVD---YQPDKLEFA-DRSITRDELKAGLHLYGTGDNDD---IKDHADW 1091  
OY 306 KRVKLKREKISIAEGSF 322  
DB 1092 DSVIEGKGNDILRGGY 1108

RESULT 12  
US-08-486-099-116  
Sequence 116, Application US/08486099  
Patent No. 6013263  
GENERAL INFORMATION:  
APPLICANT: Bolognesi, Danl P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteaway, Stephen R.  
APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF  
TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,099  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-031  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 116:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 607 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-486-099-116

Query Match 5.4%; Score 93; DB 3; Length 607;  
Best Local Similarity 21.2%; Pred. No. 0.55;  
Matches 69; Conservative 53; Mismatches 117; Indels 86; Gaps 17;

OY 43 ELARDPSSTRLDLTNVVDVYSGASGIVPEEDMVVDLGINNSVLLTPSARLOAVYKNSV 102  
DB 13 EBAEEQENLRAEG--DKRYSGRDSL-----FLVD----- 42  
OY 103 VAPAVYKSESKRYAGDTILGVRVLPFSY--SOSAMIMPFKIPFYSGSGNOFLGKGLID 161

Db 43 ASKAMEFSQSEDELTPFDMISQIOGVYISKIISDRDLAVVEYGTREKOKNSVN--FK 99  
Qy 162 NIKTMEIKVSVYSLGYEIDL-----EVLFEPMNGMEIAYSMGTLFKGKADLIMSNPN 215  
Db 100 NIYVDELDPNGAKRILELDFQKGQOKRFQDMGHGSDSL-----SEVLWVCAN 151  
Qy 216 YIPNISRRI-----IKDDVP--NYPLASSKMRFA-----FRVSKSHSSKVKNF 258  
Db 152 LFSVDVFKMSHKRIMLFTNEDNPHGNDASAKASARTKAGDLRTGTGIFLDMHLKPGCF- 210  
Qy 259 FVYKDLRVLY-DKLSVSDSIDSDISESVFKY--ETSGTES-LRKLAHETFKRY-----L 309  
Db 211 ----DISLFYRDIISIAEDEDL-----RVHFEESSKLEDLRKVRAKETRKRLSLRL 260  
Qy 310 KLRKISIAEGSFQNEVEKIESEKP 334  
Db 261 KLNKDVIISVGIV-NLVQKALKPPP 284

## RESULT 13

US-08-484-223B-116  
; Sequence 116, Application US/08484223B  
; Patent No. 6020459  
; GENERAL INFORMATION:  
; APPLICANT: Bolognesi, Daniel P.  
; APPLICANT: Matthews, Thomas J.  
; APPLICANT: Wild, Carl T.  
; APPLICANT: Barney, Shawn O.  
; APPLICANT: Lambert, Dennis M.  
; APPLICANT: Petteway, Stephen R.  
; APPLICANT: Langlois, Alphonse J.  
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF  
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
; NUMBER OF SEQUENCES: 245  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/484,223B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7872-029  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 116:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 607 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-08-484-223B-116

Query Match 5.4%; Score 93; DB 3; Length 607;  
Best Local Similarity 21.2%; Pred. No. 0.55;  
Matches 69; Conservative 53; Mismatches 117; Indels 86; Gaps 17;

Qy 43 ELARDPSTRDLTNVYVYSGASGIVKPEDMYVDLGINNWSVLLTPSARLQAYKNSV 102  
Db 13 EEAEEDQENLEASG--DYVYSGHDSLI-----FLVD----- 42  
Qy 103 VAPAVVKSSESKRYRAGDTILGVRLFPSPY-SOSSAMIMPPKIPYSGESGNOFLGKGLID 161  
Db 43 ASKAMEFSQSEDELTPFDMISQIOGVYISKIISDRDLAVVEYGTREKOKNSVN--FK 99  
Qy 162 NIKTMEIKVSVYSLGYEIDL-----EVLFEPMNGMEIAYSMGTLFKGKADLIMSNPN 215  
Db 100 NIYVDELDPNGAKRILELDFQKGQOKRFQDMGHGSDSL-----SEVLWVCAN 151  
Qy 216 YIPNISRRI-----IKDDVP--NYPLASSKMRFA-----FRVSKSHSSKVKNF 258  
Db 152 LFSVDVFKMSHKRIMLFTNEDNPHGNDASAKASARTKAGDLRTGTGIFLDMHLKPGCF- 210  
Qy 259 FVYKDLRVLY-DKLSVSDSIDSDISESVFKY--ETSGTES-LRKLAHETFKRY-----L 309  
Db 211 ----DISLFYRDIISIAEDEDL-----RVHFEESSKLEDLRKVRAKETRKRLSLRL 260  
Qy 310 KLRKISIAEGSFQNEVEKIESEKP 334  
Db 261 KLNKDVIISVGIV-NLVQKALKPPP 284

## RESULT 14

US-08-919-597-116  
; Sequence 116, Application US/08919597  
; Patent No. 6054265  
; GENERAL INFORMATION:  
; APPLICANT: Bolognesi, Daniel P.  
; APPLICANT: Matthews, Thomas J.  
; APPLICANT: Wild, Carl T.  
; APPLICANT: Barney, Shawn O.  
; APPLICANT: Lambert, Dennis M.  
; APPLICANT: Petteway, Stephen R.  
; APPLICANT: Langlois, Alphonse J.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION  
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
; NUMBER OF SEQUENCES: 273  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/919,597  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/470,896  
; FILING DATE: 06-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7872-020  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 116:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 607 amino acids  
; TYPE: amino acid

STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-919-597-116

Query Match 5.4%; Score 93; DB 3; Length 607;  
Best Local Similarity 21.2%; Pred. No. 0.55;  
Matches 69; Conservative 53; Mismatches 117; Indels 86; Gaps 17;

QY 43 ELARPPSSRLDLTNVYVYSGAGIVPEDMVVDLGINNMSVLLTPSARLQAYKNSV 102  
DB 13 EEAEEQEENLEASG--DYKYSGRDSL-----FLVD----- 42  
QY 103 VAPAVYKSKRYAGDTLLGVVLPFSY--SOSAMIMPFKIPFYSGESGNQFLGGLID 161  
DB 43 ASKAMFESSEDELPPFDMSTOCIOSVYISKIISDRDLAVFYGTEDKNSVN---FK 99  
QY 162 NIKTMEIKVSYSLGYEIDL-----EVLPEDMNGMEYAYSMGLKFKGMADLWSPN 215  
DB 100 NIYVQLDLPNPAKRLLEIDQFKGOOGKRPQDMMGHSGDYL-----SEVLWVCAN 151  
QY 216 YIPNISR-----IKDVP--NYPLASSKMRFA-----FRYSKSHSKYKNFI 258  
DB 152 LFSVDQFKSKHRKIMLTNEDNPHGNDASAKASRARTKAGDLRDTGIFLDMHLKKGPF- 210  
QY 259 FVYKDLRVLY-DKLSYSDSDIDSESVFKVY--ETSGTES-LRKLAHETPKRV-----L 309  
DB 211 ----DISLFYRDIISAEDEDL-----RVHFEESKLEDLRKVAKETRRALSRL 260  
QY 310 KLRKISIAEGSFQNFVEKIESEKP 334  
DB 261 KLNKDIIVISVGII-NLYQKALKPPP 284

RESULT 15  
US-08-475-668A-116

; Sequence 116, Application US/08475668A  
; Patent No. 6060065

; GENERAL INFORMATION:

; APPLICANT: Barney, Shawn O.  
; APPLICANT: Lambert, Dennis M.  
; APPLICANT: Peteway, Stephen R.  
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE  
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS  
; TITLE OF INVENTION: TRANSMISSION  
; NUMBER OF SEQUENCES: 211  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/475,668A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Cortuzi, Laura A.

; REGISTRATION NUMBER: 30,742

; REFERENCE/DOCKET NUMBER: 7872-026

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-9741/8864

; TELEX: 66141 PENNITE

; INFORMATION FOR SEQ ID NO: 116:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 607 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-475-668A-116

Query Match 5.4%; Score 93; DB 3; Length 607;  
Best Local Similarity 21.2%; Pred. No. 0.55;  
Matches 69; Conservative 53; Mismatches 117; Indels 86; Gaps 17;

QY 43 ELARPPSSRLDLTNVYVYSGAGIVPEDMVVDLGINNMSVLLTPSARLQAYKNSV 102  
DB 13 EEAEEQEENLEASG--DYKYSGRDSL-----FLVD----- 42  
QY 103 VAPAVYKSKRYAGDTLLGVVLPFSY--SOSAMIMPFKIPFYSGESGNQFLGGLID 161  
DB 43 ASKAMFESSEDELPPFDMSTOCIOSVYISKIISDRDLAVFYGTEDKNSVN---FK 99  
QY 162 NIKTMEIKVSYSLGYEIDL-----EVLPEDMNGMEYAYSMGLKFKGMADLWSPN 215  
DB 100 NIYVQLDLPNPAKRLLEIDQFKGOOGKRPQDMMGHSGDYL-----SEVLWVCAN 151  
QY 216 YIPNISR-----IKDVP--NYPLASSKMRFA-----FRYSKSHSKYKNFI 258  
DB 152 LFSVDQFKSKHRKIMLTNEDNPHGNDASAKASRARTKAGDLRDTGIFLDMHLKKGPF- 210  
QY 259 FVYKDLRVLY-DKLSYSDSDIDSESVFKVY--ETSGTES-LRKLAHETPKRV-----L 309  
DB 211 ----DISLFYRDIISAEDEDL-----RVHFEESKLEDLRKVAKETRRALSRL 260  
QY 310 KLRKISIAEGSFQNFVEKIESEKP 334  
DB 261 KLNKDIIVISVGII-NLYQKALKPPP 284

Search completed: March 19, 2002, 23:32:10  
Job time: 4617 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 23:32:14 ; Search time 2311.81 Seconds  
(without alignments)  
7729.973 Million cell updates/sec

Title: US-09-004-395-1

Perfect score: 1663

Sequence: 1 atgataacttttttcaaaa.....atactctgaagaataatt 1663

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estcom:\*  
5: em\_estpl:\*  
6: em\_estda:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gD\_estcl:\*  
11: gD\_estc2:\*  
12: gD\_hic:\*  
13: gD\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	98.6	5.9	1101	13	CNS00EVL
C 2	94.6	5.7	1101	13	AL069706 Drosophila
C 3	93.4	5.6	1101	13	AL069493 Drosophila
C 4	88	5.3	928	13	CNS00DXY
C 5	86.8	5.2	1101	13	CNS00E07
C 6	86.2	5.2	1029	13	CNS01ZGM
C 7	85.6	5.1	996	13	CNS00E07
C 8	85.4	5.1	1101	13	CNS00E07
C 9	85.2	5.1	1101	13	CNS00E07
C 10	84.4	5.1	759	13	CNS00E07
C 11	84	5.1	1101	13	CNS00E07
C 12	83.6	5.0	945	13	CNS00E07

C 13	83.6	5.0	1101	13	CNS0039G	AL063921 Drosophila
C 14	83.2	5.0	1043	13	CNS0145P	AL003735 Drosophila
C 15	82.8	5.0	1101	13	CNS0021J	AL061936 Drosophila
C 16	81.6	4.9	639	13	CNS03BCX	AL233458 Tetradon
C 17	81.4	4.9	910	13	CNS01G8P	AL142826 Anopheles
C 18	81	4.9	1101	13	CNS00L72	AL078714 Drosophila
C 19	80.4	4.8	1092	13	CNS020K7	AL175696 Tetradon
C 20	80	4.8	1032	13	CNS020P	AL206746 Tetradon
C 21	79.6	4.8	1201	13	CNS0167M	AL108396 Drosophila
C 22	79	4.8	843	13	CNS0091L	AL053009 Drosophila
C 23	79	4.8	1101	13	CNS002FG	AL062437 Drosophila
C 24	78.8	4.7	905	13	CNS00KXH	AL077798 Drosophila
C 25	78.6	4.7	1101	13	CNS0008H	AL063632 Drosophila
C 26	78.6	4.7	1101	13	CNS0586V	AL325912 Tetradon
C 27	78.4	4.7	1101	13	CNS00E7W	AL068447 Drosophila
C 28	78.2	4.7	609	13	CNS025K2	AL182171 Tetradon
C 29	78.2	4.7	843	13	CNS00C51	AL059666 Drosophila
C 30	77.6	4.7	987	13	CNS014P0	AL104456 Drosophila
C 31	77.6	4.7	1101	13	CNS00E07	AL069440 Drosophila
C 32	77.2	4.6	970	13	CNS0182A	AL108796 Drosophila
C 33	77.2	4.6	994	13	CNS04N0J	AL298972 Tetradon
C 34	77	4.6	836	13	CNS01100	AL099642 Drosophila
C 35	77	4.6	1027	13	CNS02T50	AL212733 Tetradon
C 36	77	4.6	1092	13	CNS020K7	AL175696 Tetradon
C 37	76.6	4.6	799	13	A2184728	A2184728 SP.1003.B
C 38	76.4	4.6	1061	13	CNS015LM	AL105604 Drosophila
C 39	76.4	4.6	1102	13	CNS014P4	AL104434 Drosophila
C 40	76.2	4.6	1101	13	CNS003B4	AL064082 Drosophila
C 41	76.2	4.6	1101	13	CNS003BD	AL064091 Drosophila
C 42	76	4.6	1131	13	CNS034P0	AL227373 Tetradon
C 43	75.8	4.6	661	13	CNS020V7	AL209807 Tetradon
C 44	75.8	4.6	668	10	AL514901	AL514901
C 45	75.8	4.6	1101	13	CNS017V2	AL108536 Drosophila

## ALIGNMENTS

RESULT 1  
CNS00EVL/c  
LOCUS  
DEFINITION  
BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION  
AL069706  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

- Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuo Oosawa and Aaron Mammosier in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).





```

DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
              BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
              fly) genomic survey sequence.
ACCESSION   AL069493
VERSION     AL069493.1  GI:4949636
KEYWORDS    GSS.
SOURCE      fruit fly.
            Drosophila melanogaster
ORGANISM    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1101)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mamoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            p1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

FEATURES             Location/Qualifiers
     source           1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone-lib="RPCI-98"
                     /clone="BACR29P01"
                     /note="end : T7"

BASE COUNT   289 a      155 c      118 g      307 t      232 others
ORIGIN
1
289 a      155 c      118 g      307 t      232 others

Query Match          5.6%: Score 93.4; DB 13; Length 1101;
Best Local Similarity 37.6%: Pred. No. 0.00014;
Matches 188; Conservative 91; Mismatches 234; Indels 4; Gaps 1;

QY 25 ttttttattcattctgaagaagatttgttgcataattaagaatttaagaagaac 84
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1040 ttttttttttdwrtgtaatrtrgtttaktktmatrarmwtmadagatgctttdgttttrt 981
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 85 aaaaattgaaagataattagattctttagtaaaagaatttcttcttaggaagcaatgag 144
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 980 ttrtkdramamwmttttkkgtwttgwtgwtttakttdrttdttadktaatttgagtgaa 921
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 145 gttttcttcttggtcttttaagtgatccaagattttatatcttttttaagaagt 204
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 920 aktrrmawmtwmaataakatttttagartrktgttrratrttdtttdakagcttttgatrgaga 861
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 205 gaatttattatgaataactc-----tcaagaagaattcatgattctttaaagaagtgat 260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 860 kagattttwtatgtaataatadtdagwmataataatataaaaraawttttatatawmaaa 801
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 261 tataatgaatccttcttgatgctgaagtcattcttgcatagtcttctaataaacaatg 320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 800 ttrtttaamaaarmdrtaaaamataatatttttaaaaaataattttttataaataatttw 741
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 321 cgccttaacatactcgttaagataaaagaataataatacttctgttgaagggtgatactta 380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 740 amwttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 681
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 381 aaggaagcaagctgataaataatttactaataaataaataaataaataaataaataatttt 440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 680 ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 621
QY 441 ataaagattatataaaggagttgtttacatgaagaaggaagcctaaagtatttat 500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 620 waaaaaaatttttaradatttaatttaamwttaamaatttttaamaatttttttttttttt 561
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 501 ttttttataaccactgtcttcttcttgctcaagagcagtgatga 547
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 560 ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 514
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
CNS00DKY 928 bp DNA GSS 04-JUN-1999
LOCUS     Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071865
VERSION     AL071865.1  GI:4948170
KEYWORDS    GSS.
SOURCE      fruit fly.
            Drosophila melanogaster
ORGANISM    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 928)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mamoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            p1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

FEATURES             Location/Qualifiers
     source           1..928
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone-lib="RPCI-98"
                     /clone="BACR27A24"
                     /note="end : T7"

BASE COUNT   262 a      70 c      84 g      321 t      191 others
ORIGIN
1
262 a      70 c      84 g      321 t      191 others

Query Match          5.3%: Score 88; DB 13; Length 928;
Best Local Similarity 32.2%: Pred. No. 0.0009;
Matches 118; Conservative 100; Mismatches 148; Indels 0; Gaps 0;

QY 161 ttatttaagatccaagattttatatcttctttaaagaatgagtttatatgtaaa 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 510 ttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 569
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 221 tcttcaagaagaattcatgattcttctaataatgagtgatgaatgaatcttgcagt 280
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 570 ttaatttttaatttttttttttttttttttttttttttttttttttttttttttttttttt 629
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 281 taaggtcaatcttcttgatgcttctaataaacaatgagccttactactcctggtaa 340
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 630 ataatatttttttttttttttttttttttttttttttttttttttttttttttttttttttt 689
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

OY	341	gataaaagaattatctcttctgtgaagggcgtatcttaagaagagcgaagtgtgataat	400
Db	650	AWTAAATAAAAAATTATTATTTTTTTTTTTTTTTTWTATWTAAATATAAATAATATMMWWATW	749
OY	401	taaatlctactaataaataataaaaaaacgaaaaattlataaagaattatataaag	460
Db	750	TWDCKNMNNNNMAMMMWWMMMAAMMMMAAMMMWWMAAAAAAAMMAAAAAA	809
OY	461	gaagtgtgattacatgaagaagcctaagaattattatltttttatccactgtt	520
Db	810	AMDDDDDDDDMMKAAKKKKKKKKKKKKKKKKKKKKTKTKTKTGAWMTTTTTTTTT	869
OY	521	ctttt	526
Db	870	TTTTTT	875
RESULT	5		
CNS00E07			
LOCUS	1101 bp	DNA	GSS
DEFINITION	Drosophila melanogaster genome survey sequence TERT3 end of BAC: BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		04-JUN-1999
ACCESSION	AL069440		
VERSION	AL069440.1	GI:4949583	
KEYWORDS	GSS.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
AUTHORS	1 (bases 1 to 1101)		
TITLE	Genoscope.		
JOURNAL	Direct Submission		
COMMENT	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .		
FEATURES			
SOURCE	1..1101	Location/Qualifiers	
	/organism="Drosophila melanogaster"		
	/db_xref="taxon:7227"		
	/clone_lib="RPCI-98"		
	/clone="BACR29P01"		
	/note="end : TERT3"		
BASE COUNT	366 a 66 c 104 g 351 t	214 others	
ORIGIN			
Query Match	5.2% Score 86.8; DB 13; Length 1101;		
Best Local Similarity	37.1% Pred. 0.0012;		
Matches 211; Conservative 84; Mismatches 273; Indels 1; Gaps 1;			
OY	15	tcaaaaaaggttttattctatctcagcaaggagattgttgcctaattgaatatla	74
Db	428	TKAATAAARFGGCGCAATATATTGAAAGATATATAAAAAARATAAATTAATTAATTA	487
OY	75	agagatgaacaaatttgaaagataattagatcttcttagtaaaagaattctttttagga	134

[illegible]

RESULT	6
LOCUS	CNS01ZGM
DEFINITION	Tetradon nigroviridis genome survey sequence pUC-ori end of clone 220112 of library G from Tetradon nigroviridis, genomic survey sequence.
ACCESSION	CNS01ZGM 1029 bp DNA
VERSION	AL174271.1 GI:7812328
KEYWORDS	GSS: genome survey sequence.
SOURCE	AL174271
ORGANISM	Tetradon nigroviridis. Tetradon nigroviridis. Tetradon nigroviridis. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE	1 (bases 1 to 1029) Roest-Crollius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Queclier,F., Saurin,W., Bernot,A. and Weissbach,J.
AUTHORS	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
TITLE	Unpublished
JOURNAL	2 (bases 1 to 1029)
REFERENCE	Roest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Queclier,F., Saurin,W. and Weissbach,J.
AUTHORS	Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence
TITLE	Unpublished
JOURNAL	3 (bases 1 to 1029)
REFERENCE	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/Tetradon">http://www.genoscope.cns.fr/Tetradon</a> .
COMMENT	Location/Qualifiers
FEATURES	



DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29M12 of RPC1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	
ACCESSION	AL069526	
VERSION	AL069526.1 GI:4949669	
KEYWORDS	GSS.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster Euarystota, Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101)	
REFERENCE	Genoscope.	
AUTHORS	Direct Submission	
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : segreif@genoscope.cns.fr	
JOURNAL	- Web : www.genoscope.cns.fr)	
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mamosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPC1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .	
FEATURES	Location/Qualifiers	
Source	1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPC1-98" /clone="BACR29M12" /note="end : 77"	
BASE COUNT	249 a 72 c 234 g 360 t 186 others	
ORIGIN		
Query Match	5.1%; Score 85.4; DB 13; Length 1101;	
Best Local Similarity	39.3%; Pred. No. 0.0019;	
Matches 203; Conservative	56; Mismatches 257; Indels 0; Gaps 0.	
2	tgatactcttttcaaaaaggttttatttattcttagcaagagattgttgctaa 61	
552	KKNITTTTTTTTTTTTTTTTTTTTTTTCTGNGAGCGRRIRAWTTMKIKKTYT 611	
62	ttaagatalttaagatgaagaatltgaagaataatttagatcttttgataaaga 121	
612	TKTKTTTTTTTTTTTTTTATKATRAGAAAAMARAGAGWGCGGGCGTGCTKTKTTTTT 671	
122	ttctctttagaagaatgagggcttctcttggtgtttatttaagtgtatcaagatt 181	
672	TTTTTTTTTAAARANAAMWTTTTTTTATTTTTTTTTTTTTTTTAAATATKAGAMGKCT 731	
182	tttatatctttttaaagaatgaggttattatgttaaatcttccaagaatttatga 241	
732	TTTTTTTTTAAWMAAAATTTAAAAAAWMAAAWMAAAATTTTTTTTAAWATWAAKTAT 791	
242	ttctttaaataatgltgatttaagaatctttagatgtgaagtcattcttgctat 301	
792	GATAAATATMAATTTTTTTTATATWAAAMWTTWTTTTTTTATATTAATTAATAAAMAAAAA 851	
302	gtctttaataaacaatgcgttctaactatccgtgaagataaagaagattatattct 361	
852	AAAAATTTTATATATKATTTTATTTTATATTTTAAWMAAAATWAAAWATTTTTTTTTTTT 911	
362	tgttgaagggtatctcaagaagcaagttgataaataatttactaataaataaata 421	

Db	912	TWTWTAATWATTTTKNNATFWMAATAAATTWTTTATNAAAMWATTTTAATGAAAAAMWAATTT	971
Oy	422	attaaagaacgaatcttataaagattatatacgaagggttgattacaagaag	481
Db	972	TWATTTATATWMAATTTTTTTAAATTAATTAATTAAMWAATMTWMTTNAAAAMAAAAA	1031
Oy	482	aagctaagaagtattatttttcttataccact	517
Db	1032	AWAANTAAAWMAWWMAATTTAAATTAATTAATTT	1067
RESULT	9		
CNS0021J			
LOCUS	CNS0021J	1101 bp	DNA GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TEI3 end of BAC # BACR05N11 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL061936		
VERSION	AL061936.1	GI:4940214	
KEYWORDS	GSS.		
SOURCE	Fruit fly.		
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr Web : www.genoscope.cns.fr)		
REFERENCE	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named Rpci-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.		
FEATURES			
source	Location/Qualifiers 1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_1lb="Rpci-98" /clone="BACR05N11" /note="end : TEI3"		
BASE COUNT	631 a	7 c	28 g 289 t 146 others
ORIGIN			
Query Match	5.1%; Score 85.2;	DB 13;	Length 1101;
Best Local Similarity	41.2%;	Pred. No. 0.0021;	
Matches	219;	Conservative 52;	Mismatches 260; Indels 1; Caps 1;
Oy	60	aattaagatatttaagsagatgacaaatttgaaagataattgatcccttagtaaaa	119
Db	450	AAAN	509
Oy	120	gatttcctttaaggaagaagggggttccttccttggtttttaaagatccaaga	179
Db	510	AAAAATATAATTTATTTTWTWTAATTAAATTAATTTTTTTTTTTTTTTTWTWTAATTTTAA	569
Oy	180	ttttaatactcttttaagaagaatggattataatgcaatccttccaagaacttat	239
Db	570	TTTTTAANAAMAATTTAATAAANAWTWTITTAATTTTAATWTAAMWAAAAAAATTTTA	629





DB	411	AAAMDNBNMAAAAAAANNNNNGATT	367
RESULT	13	:     :	
CNS0039G/C		CNS0039G	1101 bp . DNA GSS 03-JUN-1999
LOCUS		Drosophila melanogaster genome survey sequence TETJ3 end of BAC #	
DEFINITION		BACR08K10 of RPrCt-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	
ACCESSION		AL063921	
VERSION		AL063921.1	GI:4941778
KEYWORDS		GSS.	
SOURCE		fruit fly.	
ORGANISM		Drosophila melanogaster.	
REFERENCE		Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
AUTHORS		Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
TITLE		Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
JOURNAL		1 (bases 1 to 1101)	
COMMENT		Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPrCt-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.	
FEATURES		Location/Qualifiers	
source		1..1101	
		/organism="Drosophila melanogaster"	
		/db_xref="taxon:7227"	
		/clone_lib="RPrCt-98"	
		/clone="BACR08K10"	
		/note="end : TETJ3"	
BASE COUNT		201 a 64 c 131 g 202 t 503 others	
ORIGIN			
Query Match		5.0%, Score 83.6; DB J3; Length 1101;	
Best Local Similarity		14.3%; Pred.No.0.0035;	
Matches		70; Conservative 265; Mismatches 150; Indels 3; Gaps 1;	
OY	25	tttttatattcatctcgtcaaggaagtcttgctgccaatttaagaatatlaagagatgaac	84
Db	1041	mmmmmwamwmdtmwmwmatraktdtmtwtmrtrmrdwmaordgagckrdadattdag	982
OY	85	aaaattgaagataatlagactctttaagtaaagaattcttcittagaagaactaggg	144
Db	961	agrrbggrkrrdkrrrkodddkkcgkkrakkaakawmkrmwdmddmwdmwdakdkk	922
OY	145	ggttctctcttggtttatctaagtgatccaagaatttatactcttttcaaagaatg	204
Db	921	addddgcagdkgdgdgkgnadddotgctkoddokdkdmddmkacgtwgdatwmaadtwwm	862
OY	205	gagttatatagtaaactcttcaagaagaatttatgatctctttaataatggtgatata	264
Db	861	gwadadnwmt---wdaaadwmadrwdmamkwmdamwagartadrpmgdragrkgar	805
OY	265	atgaactcttgatgttgtaagtgcaactctttgtgcatagtcttaataaaaaaatggct	324
Db	804	krrdkrraddkdaadddrdaattwttttttttttttttttttttttttttttttttttt	745

Qy	325	ttaactatccgtgaagataaaaaagattatattctgttgaagggtgtactcaaaag	384
Db	744	DHRRGTAGTGRKKRRRTWKRRKKRRDTRWMDADDDTARDDRRRRGDDGAGCKTKGRKR	685
Qy	385	agcaaatgtcaataaattactaataaaaaataaataaaacgaaaaatttataa	444
Db	684	RRDRDATTMDRDTAAWADAAMWTTTDTDDMDCKRDRRRKCGARRRRRTTARAAMDMWTWKA	625
Qy	445	aagattatataataaggagtggttctacatagaaagaaagctaagaattattttt	504
Db	624	WDMAWMDWKTDRADRDWADDTWTDARADRDMAKARARARRDRARARADRRWTTGCK	565
Qy	505	ttattat	512
Db	564	TTTTATWTT	557
RESULT	14		
LOCUS	CNS0145P/C		
DEFINITION	CNS0145P 1043 bp DNA GSS 26-JUL-1999		
ACCESSION	CNS0145P		
VERSION	AL103735		
KEYWORDS	AL103735.1 GI:5615346		
SOURCE	GSS		
ORGANISM	fruit fly.		
REFERENCE	Plasmid Drosophila melanogaster		
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
JOURNAL	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
COMMENT	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
	1 (bases 1 to 1043)		
	Genoscope.		
	Direct Submission		
	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage		
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
	- Web : www.genoscope.cns.fr)		
	Determination of this BAC-end sequence was carried out as part of a		
	collaboration with the European Drosophila Genome Project (EDGP) -		
	http://www.edgp.ebi.ac.uk.. This Drosophila melanogaster BAC		
	library (Dros BAC) was made by Alain Billaud at CEPH (Centre		
	d'Etude du Polymorphisme Humain) with funding provided by a MRC		
	project grant. The DNA was prepared from embryos by Alain Bucheton		
	and Genevieve Payan. It has been constructed in the vector		
	pBelobAC11.		
FEATURES	Location/Qualifiers		
Source	1..1043		
	/organism="Drosophila melanogaster"		
	/plasmid="pBelobAC11"		
	/db_xref="taxon:7227"		
	/clone_lib="DrosBAC"		
	/clone="BACN11G11"		
	/note="end : 77"		
BASE COUNT	277 a 96 c 121 g 382 t 167 others		
ORIGIN			
Query Match	5.0%; Score 83.2; DB 13; Length 1043;		
Best Local Similarity	38.9%; Pred. No. 0.0041;		
Matches 165; Conservative 72; Mismatches 182; Indels 5; Gaps 1;			
Qy	140	tgaagggttcttcttggtgggttttataagatccaagatttatacttctttaa	199
Db	955	KKKKKKKTTTTTTTTTTTTTTTTTTTATATATTTTTTDRKTTWTDATTTGKATTTT	900
Qy	200	gaatgagtttataatgtaaacctttccaagaaatattatgattcttctaataatg	259
Db	899	RTDTRKTTTTTTTTTTTTTTTTTTT-----TTTTTTTTTTTTTTTWTWTATWT	845
Qy	260	tataatgaatcttttgatgtaagtcacatcttttgctagcttctaataaacaac	319
Db	844	WATATATWTARAGRAADDAAAAAATTTTWTWMMWMMWTTTMMWMMWTTTWTW	785



QY	320	gcgccttaacatccctgtaagacaacaaagattatctcttgtaagggtgctct	379
Db	784	TTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATNTNMAATATATAATTWA	723
QY	380	aaaggagcaagctgtgataaattlaaatttactaataaabaattaaacgaacatt	439
Db	724	AAATATATATATAATATATAATATAATATAATATATATATAATATAATATAATATA	665
QY	440	tataaagattatataaaggagtggtttacatgaagaagaaagcttaaaagattta	499
Db	664	WAWWMAAAWMTATATTTTWWMAWMTAAWMAWMAWMAAAWMTATATATTTTWWA	605
QY	500	ttttttattatccactgctcttttgcctcaagagacgtgattagacagaggctct	555
Db	604	TATATAAAAATATATATAKTTTTTTTATATWMTATWMTATATATATWMAAAAAAGAGA	549
QY	560	aaaa 563	
Db	544	AAAA 541	

RESULT	15
CNS00211/c	
LOCUS	
DEFINITION	CNS00211 1101 bp DNA GSS 03-JUN-1999
DESCRIPTION	Drosophila melanogaster genome survey sequence Tetr3 end of BAC # BACR05N11 of Rpct-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	U00002

ACCESSION	AL061936	GI:4940214
VERSION	AL061936.1	
KEYWORDS	GSS.	
SOURCE	fruit fly.	
ORGANISM	Drosophila	melanogaster

REFERENCE 1 (bases 1 to 1101)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

**COMMENT**

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see [http://www.fruitfly.org/The\\_BDGP/Drosophila\\_melanogaster\\_BAC\\_library](http://www.fruitfly.org/The_BDGP/Drosophila_melanogaster_BAC_library) was prepared by Kazutoyo Oseegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES	source	location/Qualifiers
BASE COUNT	631 a	289 t
ORIGIN	7 c	146 others
		1. .1101
		/dbxref="Drosophila melanogaster"
		/obj_xref="taxon:7227"
		/clone_lib="RPCI-98"
		/clone="BACRO5N11"
		/note="end : TE13"

Query Match	5.0%;	Score 82.8;	DB 13;	Length 1101;
Best Local Similarity	42.0%;	Pred. No. 0.0046;		
Matches 215;	Conservative 46;	Mismatches 246;	Indels 3;	Gaps 1;
0y	18	aaaaagttttattatctcatctcagcaggatttgttgcaatttaagatatattaaga	77	

[illegible]

```
Search completed: March 20, 2002, 00:45:13
Job time: 4379 sec
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 23:42:14 ; Search time 116.86 Seconds  
(without alignments)  
3222.941 Million cell updates/sec

Title: US-09-004-395-1

Perfect score: 1663  
Sequence: 1 atgataatcttttcaaaa.....atatccttgaagaatatt 1663

Scoring table: IDENTITY-NUC  
Gap 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents.NA:\*  
1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PCRTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	4.7	19124	2	US-08-487-826B-13
2	63	3.8	665	2	US-08-883-795A-36
3	62.4	3.8	837	4	US-08-998-416-288
4	62.4	3.8	837	4	US-08-998-416-288
5	61.8	3.7	615	4	US-08-998-416-186
6	61.4	3.7	636	4	US-08-998-416-1137
7	59.6	3.6	615	4	US-08-998-416-186
8	59.2	3.6	636	4	US-08-998-416-1137
9	56	3.4	8920	2	US-08-446-855A-1
10	56	3.4	8920	2	US-09-150-741-1
11	55.4	3.3	4673	1	US-07-638-431-1
12	55.4	3.3	4673	1	PCR-US92-00018-1
13	55	3.3	19124	2	US-08-487-826B-13
14	54.4	3.3	665	2	US-08-883-795A-36
15	54	3.2	4818	2	US-08-817-925-27
16	53	3.2	6243	2	US-09-056-075-1
17	51.4	3.1	3600	1	US-08-894-731-1
18	51.2	3.1	1511	1	US-07-991-867B-8
19	51.2	3.1	1511	1	US-08-107-755A-8
20	51.2	3.1	1511	2	US-08-544-332-8
21	50.8	3.1	722	4	US-08-998-416-780
22	50.2	3.0	782	4	US-09-007-119-15
23	50.2	3.0	4810	3	US-08-852-629-11
24	50.2	3.0	4838	3	US-08-852-629-15
25	50	3.0	642	1	US-08-764-100-13
26	50	3.0	665	4	US-08-998-416-937
27	50	3.0	724	4	US-09-007-119-8

28	50	3.0	3000	1	US-08-764-100-9	Sequence 9, Appl
29	50	3.0	6243	2	US-09-056-075-1	Sequence 1, Appl
30	49.8	3.0	3095	6	523158-1	Patent No. 523158
31	49.8	3.0	51952	3	US-08-947-823-1	Sequence 1, Appl
32	49.4	3.0	727	4	US-08-998-416-1011	Sequence 1011, Ap
33	49	2.9	660	1	US-07-991-867B-32	Sequence 32, Appl
34	49	2.9	660	1	US-08-107-755A-32	Sequence 32, Appl
35	49	2.9	660	2	US-08-544-332-32	Sequence 32, Appl
36	49	2.9	1511	1	US-07-991-867B-8	Sequence 8, Appl
37	49	2.9	1511	1	US-08-107-755A-8	Sequence 8, Appl
38	49	2.9	1511	2	US-08-544-332-8	Sequence 8, Appl
39	49	2.9	1654	3	US-08-913-842-20	Sequence 20, Appl
40	49	2.9	1875	3	US-08-913-842-1	Sequence 1, Appl
41	48.8	2.9	658	4	US-08-998-416-595	Sequence 595, App
42	48.8	2.9	4431	4	US-09-532-803-8	Sequence 8, Appl
43	48	2.9	5852	1	US-07-867-106-2	Sequence 2, Appl
44	47.8	2.9	319	1	US-07-593-657-14	Sequence 14, Appl
45	47.6	2.9	1172	1	US-07-945-288-9	Sequence 9, Appl

## ALIGNMENTS

RESULT 1  
US-08-487-826B-13/c  
Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knodde Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487, 826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned 29,655  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13  
Query Match 4.7%; Score 78; DB 2; Length 19124;

Best Local Similarity 48.1%; Pred. No. 3.7e-07;  
Matches 222; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

```
OY 1 atgataacatttttcaaaaaagtttttatttatttcttctagcaaggatttctgcta 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15832 ATATATTTTATTTTAAATAATTTTATTTTATTTTATTTTATTTTATTTTAC 15773
OY 61 atttaagatatttaagagatgaacaaatttgaaagaattagacttttagtaaa 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15772 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTCT 15713
OY 121 atttcttctgaagcaagaggggttctcttgggttttatttaagtgatcaag 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15712 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTCT 15653
OY 181 ttcttatctttttttaaagaatgaggtttattatgtaaaccttccaagagaatt 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15652 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 15593
OY 241 attctttaaatgagtgattataatgaactcttgatgtaaggccaactcttctg 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15592 ATAAATCTTTTTCATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15533
OY 301 tgcctttaataaaacaaatgcgctttaaacttcctggaagataaaagaattat 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15532 TTTTAAAAATTTTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 15473
OY 361 ttgttgaagggtgatctcttaagaagcaagttgataaatttactaataaata 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15472 TATTTTCATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 15413
OY 421 aattaaaaacgaanaatttataaagaattatataaaga 462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15412 AATCATATATATATAAATATATATATAAACAACATACATACATACAT 15371
```

## RESULT 2

```
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcive, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Michelle
; REGISTRATION NUMBER: 40,261
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: Rh 32
; US-08-883-795A-36
```

Query Match 3.8%; Score 63; DB 2; Length 665;  
Best Local Similarity 47.8%; Pred. No. 0.00018;  
Matches 249; Conservative 0; Mismatches 265; Indels 7; Gaps 2;

```
OY 6 aatcttttcaaaaaagtttttatttatttcttcttctagcaaggagttgttcaatta 65
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 33 AATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 92
OY 66 agatatttaagagatgaacaaatttgaaagataatttagacttttagtaaaagatt 125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93 AAATATTTA--TAATTAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 150
OY 126 cttaagagaagaatgaggggttctcttgggttttattatcaagtgatcaagattta 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 TAAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 210
OY 186 taticctttttaaagaatgaggtttattatgtaaaccttccaagaaatttatgattc 245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 TAAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 270
OY 246 tttaataatgtagtataatgaatccttcttgatglttaaggtcaactcttctgctag 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 TAAATGTTTAAATTA-----AATATTTTATTTTATTTTATTTTATTTTATTT 325
OY 306 ttaataaaacaatgcgctttaaactatcctgtaagaataaaaaagattattatct 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 326 ATAAATTAAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 385
OY 366 gaagggtgactcttaagaagcaagcttgataaatttaacttaataaaataaata 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 386 ATAAATTAAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 445
OY 426 aaaaacgaanaatttataaagaattatataaagaaggttggttacaagaaagaag 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 446 ATAAATTAATTTTATTAAGTATTTTATTAATTAATTAATTAATTAATTAATTA 505
OY 486 cttaaaagtatttatttttattatcaccactgttcttt 526
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 506 TTACATATTTTATTAATTAAGTATTTTATTAATTAATTAATTAATTAATTT 546
```

## RESULT 3

```
US-08-998-416-288
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebschunig, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264tn Carolina
; COUNTRY: USA
; ZIP: 27709
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ. ID NO.: 288:
SEQUENCE CHARACTERISTICS:
LENGTH: 837 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: PAG1241RP
US-08-998-416-288
```

Query Match 3.8%; Score 62.4; DB 4; Length 837;

Best Local Similarity 45.1%; Pred. No. 0.00025; Matches 231; Conservative 0; Mismatches 281; Indels 0; Gaps 0;

```
OY 1 atgataatcttttcaaaaagggttttattcattcctagcaaggagtgtgcta 60
DB 83 ATTATATCATATTTAATAAATTATTTGATTTATTAATCTATATATTAATT 142
OY 61 attaagatacttaagagatgaacaaatttgaagataattagacttttaglaaag 120
DB 143 ATATATATTACTTAATTCATCATTTATTTATTTATATTAATAAATAATTTAA 202
OY 121 attcttcttaggaagcaatlgagggtttcttcttggttttataagtgatcaagat 180
DB 203 ATGAATACCTAATTAAGTCTATGTTCAAAATTTTAAATTGATTAAATAATTAAGATAT 262
OY 181 ttctaatctcttttaagaagtgggttataatgaatcaatcttcaagagaattatcg 240
DB 263 TATTTATTTCTTTAATAAATTATTAATAATGATTTATCAATTAATTAATTTATTTAA 322
OY 241 attcttctaataatggatgataatgaatcctttagatgtaagtcacatcttttgcta 300
DB 323 TGTATTATTAATAATATATTTTATTTATTAATAAGATTTAATTTAATAATATGTA 382
OY 301 tgccttctaataaacaacgcttaactatccctgtaagataaataaagatataatc 360
DB 383 ATATATATTATTTATTAATATCTATTTTATTAATAATTAATTAATTAATTTAA 442
OY 361 tcttgaagggtgcatcttaagaagcaagtgataaacttctaacttaataaact 420
DB 443 TCTTTTATTAAGAAATTAATTAATAATTTTAACCTTAATTTCTTATTTATTTATTT 502
OY 421 aattcaaaaacgaaacttctaataaagattatataaaggagtggttacaagaag 480
DB 503 TATATTTATTTAATAATTAATTAATTTCAATTTATTTATTTAATAATTAATTAATA 562
OY 481 gaagcctaaagatattatttttattatatt 512
DB 563 TTTAATTAATATTTATTCATTTATTTAATTAAT 594
```

RESULT 4

US-08-998-416-288/c  
Sequence 288, Application US/08998416  
Patent No. 6239264

GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Redischung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII  
NUMBER OF SEQUENCES: 1152

CORRESPONDENCE ADDRESSES:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/998,416

FILING DATE: 24-DEC-1997

CLASSIFICATION: 435

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: CH 0016/97

FILING DATE: 31-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689

INFORMATION FOR SEQ. ID NO.: 288:

SEQUENCE CHARACTERISTICS:

LENGTH: 837 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORGANISM: PAG1241RP

US-08-998-416-288

Query Match 3.8%; Score 62.4; DB 4; Length 837;

Best Local Similarity 52.2%; Pred. No. 0.00025; Matches 214; Conservative 0; Mismatches 186; Indels 10; Gaps 3;

```
OY 106 atctttagtaagaattcttctttaggaagcaatgaagggtttcttcttggtttat 165
DB 656 ATATATATAGTTCGATCCCTTTATTAAGTATTTTAACTACATTCCTTAATAATATTTAT 597
OY 166 taagtgatcaagaatttatcttctttagaagaatgggttataatgaatcctt 225
DB 596 TAAATTAATAAATGAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 537
OY 226 caagagaatttgaatcttcttctaataatgtagataatgaatcctttagatgaag 285
DB 536 AATTAATAATGAATAA-TAATTTATTTAAATTAATAATTAATAATTAATAAGAAATTAAG 479
OY 286 tcaatcttcttgc-----atgcttctaataaacaacatgcttctaactatcctgtaag 341
DB 478 TTAATAATTAATTTAATAATAATTTCTTAATAAAGATTAATAATTAATAATTAATTAAT 419
OY 342 ataaaaagaattattcttcttcttgaagggtgatactttaaaggagaagtgataaatt 401
```



TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1692RP  
US-08-998-416-1137

Query Match 3.7%; Score 61.4; DB 4; Length 636;  
Best Local Similarity 46.3%; Pred. No. 0.00038;  
Matches 243; Conservative 0; Mismatches 276; Indels 6; Gaps 1;

OY 1 atgataatcctttccaaaaaggcttttatttccattcttagcaaggagcttggtcta 60  
DB 83 ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 142  
OY 61 atttaagatatttaagagatgacaaatttgaagaataattagatctttagtaaaag 120  
DB 143 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 202  
OY 121 atttcttcttgaagagcaatgagggcttcttcttggttcttataagtgatcaagat 180  
DB 203 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 262  
OY 181 ttttataatcttctttaaagaatgagttatataatccttcaagaagaat 234  
DB 263 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 322  
OY 235 tttaagatcttataatgagtgatataagaaatcttcttgatgagcaatctt 294  
DB 323 TTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 382  
OY 295 tlycatalcctttaaacaacacgccttaactatcctgtaagataaaagatga 354  
DB 383 ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 442  
OY 355 ttatctcgttgaaggggtcatcttaaaggagcaagtgtaaaattacttaac 414  
DB 443 TCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 502  
OY 415 aaaaaaacttaaaaaagcaaaattctataaagattatataaagaagctggttcaac 474  
DB 503 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 562  
OY 475 gaaaaagaaagctaaagcatttattcttcttataatccactgt 519  
DB 563 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 607

RESULT 7  
US-08-998-416-186/C  
Sequence 186, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jürgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Reibischung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSPITII  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/C61976  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ. ID NO.: 186:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 615 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1074RP  
US-08-998-416-186

Query Match 3.6%; Score 59.6; DB 4; Length 615;  
Best Local Similarity 49.7%; Pred. No. 0.00085;  
Matches 181; Conservative 0; Mismatches 179; Indels 4; Gaps 1;

OY 148 ttctcttggcttttataagtgatcagaatttataatcttctttaaagaatgag 207  
DB 612 TTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 553  
OY 208 ttatataatgaatccttcaagaatcttataatcctttaaataatgagatcaag 267  
DB 552 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 493  
OY 268 aatccttgatgtaagcgaatcttctgataatcttataaacaatgagcgtt 327  
DB 492 AATAAGAAATTAAGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 433  
OY 328 actatcctgtaagataaaagatattatcttctgtaaggggtatcttaaaagagc 387  
DB 432 TAAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 373  
OY 388 aaagtgtataaatttaattcttaataaataaataaataaataaataaataaataa 447  
DB 372 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 317  
OY 448 atttataatgaaggttggttcaatgaaagaaagaaagaaagaaagaaagaaagaa 507  
DB 316 AATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 257  
OY 508 acta 511  
DB 256 AATA 253

RESULT 8  
US-08-998-416-1137/C  
Sequence 1137, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jürgen





Dd	434	ATATATTAATTAATTCGATTCATATTAAATGGAAATATATTAAAGTTTATATTTCAGT	375
Qy	368	agggcgtatcctaaaggagcgaagctgataaatlaaatttccactcaaaaaaataatlaa	427
Dd	374	AATTTTATTAACCAATGACACCAAAACATATATATATATATATATATATATATAT	315
Qy	428	aaacgaaaaattcctaaagaattacataaagagcttggtttcacatgaaagaaagct	487
Dd	314	ATATATATATAAATTAACCTTAATGATTTGTTATTAATATTAAGAAAAA	255
Qy	488	aaagatattctatctttctccta	511
Dd	254	AAAGCTTTTATCTATCTATTA	231

```

RESULT 10
US-09-150-741-1/c
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PLS380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 8920
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-150-741-1

```

Query Match	3.4%;	Score 56;	DB 4;	Length 8920;
Best Local Similarity	44.4%;	Pred. No. 0.0074;		
Matches 224;	Conservative 0;	Mismatches 280;	Indels 0;	Gaps 0

OY	8	tcctcttccaaaaggtttcttcttcattctcgaagaggtttgctcaattaa	67
Db	734	TTTTTTTTTATACATTTATATATTTATGTTTTAAAAATTTATAAATTTTCATATACAACTTC	675
OY	68	atattcgaagatgcacaaattcgaagaatcttgatctcttgaaagtcttc	127
Db	674	ATTTTTCATATGTAATTTTTTTTTTTCTTTTTTTTTTTTTTTTTTTTTTAA	615
OY	128	tttgaagacgaagggtttctcttggttttcaatgagatcaagatttata	187
Db	614	ATTAGTACAATTACTATTTTATTAACATACGAAAAAATAATTAATGAATTAATTAATA	555
OY	188	ttctcttcttlaaagaatgagttcttatgtcaatccttcaagaaatttatgacttc	247
Db	554	AATAAATTTATTAATAATAAATTATAGCAACCAATATTGGGACTATTATATATGTGTA	495
OY	248	caataatgctgattataatgaactcttgatggttaaggctcaactcttttgctatgcttc	307
Db	494	TAAATATATAGATGCTTTATTATATATAGCAAGTGTAACCTATATTAAATGTACACATCAA	435
OY	308	aataaaacaaatgcgttttaactaacccttgtaagataaaaagattatacttcgttga	367
Db	434	AATATTAATTAATTTGATTCATATTAAATTTGGAATAATTTAATAGCTTTTATATTTCAGT	375
OY	368	aggttgctacttcaagaagcaagtgtatataaattactactaataaataataaa	427
Db	374	AATTTTATATAAACATGACACACAAACATTTATATTTATTTATATATATATATATAT	315

```
Oy 428 aaacgaacatttttaaaagcttcatatataaagagcttggttttcacgaaaggaagc 487
Db 314 ATATATATATATAAATACCTTAAATGATATGTTAATATAATATAAGAAAAA 255
Oy 488 aaagatcttattcttttata 511
Db 254 AAAAGCTTTTATCATGATGTAATTA 231
```

```

11      RESULT
12      US-07-638-431-1/c
13      Sequence 1, Application US/07638431
14      Patent No. 5198535
15      GENERAL INFORMATION:
16      APPLICANT: Hoffman, Stephen L.
17      APPLICANT: Charoenvit, Yupin
18      APPLICANT: Hedstrom, Richard
19      APPLICANT: Khusmith, Srisin
20      APPLICANT: Rogers IV, William O.
21      TITLE OF INVENTION: Protective malaria sporozoite surface protein
22      TITLE OF INVENTION: Immunogen and gene
23      NUMBER OF SEQUENCES: 2
24      CORRESPONDENCE ADDRESS:
25      ADDRESSEE: A. David Spevack
26      STREET: NMRDC Building 1 T-12 National Naval
27      STREET: Medical Center
28      CITY: Bethesda
29      STATE: MD
30      COUNTRY: USA
31      ZIP: 20814-5044
32      COMPUTER READABLE FORM:
33      MEDIUM TYPE: Floppy disk
34      COMPUTER: IBM PC compatible
35      OPERATING SYSTEM: PC-DOS/MS-DOS
36      SOFTWARE: Patentln Release #1.24
37      CURRENT APPLICATION DATA:
38      APPLICATION NUMBER: US/07/638.431
39      FILING DATE: 19910110
40      CLASSIFICATION: 424
41      ATTORNEY/AGENT INFORMATION:
42      NAME: Spevack, Avrom D.
43      TELECOMMUNICATION INFORMATION:
44      TELEPHONE: (301) 295-6759
45      TELEFAX: (301) 295-4033
46      INFORMATION FOR SEQ ID NO: 1:
47      SEQUENCE CHARACTERISTICS:
48      LENGTH: 4673 base pairs
49      TYPE: NUCLEIC ACID
50      STRANDEDNESS: double
51      TOPOLOGY: linear
52      MOLECULE TYPE: DNA (genomic)
53      HYPOTHETICAL: N
54      ANTI-SENSE: N
55      ORIGINAL SOURCE:
56      ORGANISM: Plasmodium yoelii1
57      STRAIN: 17X(NL)
58      DEVELOPMENTAL STAGE: erythrocytic stage
59      TISSUE TYPE: Blood
60      CELL TYPE: erythrocytic stage
61      IMMEDIATE SOURCE:
62      LIBRARY: Py-lambdaag11-2-7 kb genomic expression
63      CLONE: Py10.1111
64      FEATURE:
65      NAME/KEY: CDS
66      LOCATION: 718..3195
67      OTHER INFORMATION:
68      US-07-638-431-1

```

Query Match	3.38;	Score 55.4;	DB 1;	Length 4673;
Best Local Similarity	45.78;	Pred. No. 0.0086;		
Matches 229;	Conservative 0;	Mismatches 271;	Indels 1;	Gaps 1;

```
OY 1 atgataccttttcaaaaagggttttatttcatctcgaaggatctgtccta 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3910 ATCATGAGCCAACTTGTGATATCTCTTTTAAATCCATTCATCTTTTTCACAC 3851
OY 61 attaagatattgaagatgaacaaaattgaaagataattagatcctttagaaag 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3850 TTTTACTGCTATAAACTTAAGCACTATTTATGTTTGTAAATTTTAAATATCA 3791
OY 121 atttcttttagaagaagaggggttcttcttgggttttcttaagatcatcaagat 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3790 CATTTTGTATTTCTTTTATCGATTAATGTGCGATTTGTCTATTTTATGGAAT 3731
OY 181 tttaatactttttaaagaatgagtttaatacttaaccttcaagagaattatg 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3730 TCATTTATATGATATATCATCTTTTGTATTTCATATATTTTTGAATAATAGTAAA 3671
OY 241 attccttaataatgagatataatgaatcctttagtgaaggcaatcctttgtcta 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3670 TACCGTGAATATACAACTTAATAATGTATTAACCTTTATAGCTTTTATTTT 3611
OY 301 tgtcttaataaaaacaatgcgcttaactatacctgtgaagataaaaagtattatc 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3610 TTTTATTTATTTAAGATTAATGTGAACCATTTAAATATTTGAGTATTAATATATAT 3551
OY 361 ttgtgaagggtgatct-aaagagcaaaagttgataaataatttactaaataaa 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3550 ATTATACGACACATTTAGTTTAAAAAAAATGTTAAAAAAAATCGTTAAAAAAAATA 3491
OY 420 taattaaaaaagaaaatttataaagattatataaagagtggttcaatgaaa 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3490 TGAATAATTAATGATATATTAATTAATGAATGAATGAATGAATGAATGAATGAAT 3431
OY 480 ggaagcctaaagattttat 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3430 GTTAATTTAAATAACTTAAT 3410
```

## RESULT 12

PCT-US92-00018-1/c

Sequence 1, Application PC/70US9200018

GENERAL INFORMATION:

APPLICANT: Hoffman, Stephen L.

APPLICANT: Charoenvit, Yupin

APPLICANT: Hedsstrom, Richard

APPLICANT: Khumsmith, Srisin

APPLICANT: Rogers IV, William O.

TITLE OF INVENTION: Protective malaria sporozoite surface protein

TITLE OF INVENTION: Immunogen and gene encoding

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: A. David Spevack

STREET: NMRC Building 1 T-12 National Naval

CITY: Bethesda

STATE: MD

COUNTRY: USA

ZIP: 20814-5044

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/00018

FILING DATE: 19920103

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Spevack, Avram D.

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 295-6759

TELEFAX: (301) 295-4033

INFORMATION FOR SEQ ID NO: 1:

```
SEQUENCE CHARACTERISTICS:
LENGTH: 4673 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
ORIGINAL SOURCE:
ORGANISM: Plasmodium yoelii
STRAIN: 17X(NL)
DEVELOPMENTAL STAGE: erythrocytic stage
TISSUE TYPE: Blood
CELL TYPE: erythrocytic stage
IMMEDIATE SOURCE:
LIBRARY: Py-lambda gtl1-2-7 kb genomic expression
CLONE: Py10.1111
FEATURE:
NAME/KEY: CDS
LOCATION: 718..3195
OTHER INFORMATION:
PCT-US92-00018-1
```

Query Match 3.38; Score 55.4; DB 5; Length 4673;  
Best local similarity 45.7%; Pred. No. 0.0086;  
Matches 229; Conservative 0; Mismatches 271; Indels 1; Gaps 1;

```
OY 1 atgataccttttcaaaaagggttttatttcatctcgaaggatctgtccta 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3910 ATCATGAGCCAACTTGTGATATCTCTTTTAAATCCATTCATCTTTTTCACAC 3851
OY 61 attaagatattgaagatgaacaaaattgaaagataattagatcctttagaaag 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3850 TTTTACTGCTATAAACTTAAGCACTATTTATGTTTGTAAATTTTAAATATCA 3791
OY 121 atttcttttagaagaagaggggttcttcttgggttttcttaagatcatcaagat 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3790 CATTTTGTATTTCTTTTATCGATTAATGTGCGATTTGTCTATTTTATGGAAT 3731
OY 181 tttaatactttttaaagaatgagtttaatacttaaccttcaagagaattatg 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3730 TCATTTATATGATATATCATCTTTTGTATTTCATATATTTTTGAATAATAGTAAA 3671
OY 241 attccttaataatgagatataatgaatcctttagtgaaggcaatcctttgtcta 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3670 TACCGTGAATATACAACTTAATAATGTATTAACCTTTATAGCTTTTATTTT 3611
OY 301 tgtcttaataaaaacaatgcgcttaactatacctgtgaagataaaaagtattatc 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3610 TTTTATTTATTTAAGATTAATGTGAACCATTTAAATATTTGAGTATTAATATATAT 3551
OY 361 ttgtgaagggtgatct-aaagagcaaaagttgataaataatttactaaataaa 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3550 ATTATACGACACATTTAGTTTAAAAAAAATGTTAAAAAAAATCGTTAAAAAAAATA 3491
OY 420 taattaaaaaagaaaatttataaagattatataaagagtggttcaatgaaa 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3490 TGAATAATTAATGATATATTAATTAATGAATGAATGAATGAATGAATGAATGAAT 3431
OY 480 ggaagcctaaagattttat 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3430 GTTAATTTAAATAACTTAAT 3410
```

## RESULT 13

US-08-487-826B-13

Sequence 13, Application US/08487826B

Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 3.3%; Score 55; DB 2; Length 19124;  
Best Local Similarity 47.8%; Pred. No. 0.013;  
Matches 189; Conservative 0; Mismatches 205; Indels 1; Gaps 1;  
OY 159 tttcttaagtgatccaagattcttctttaaagaatggattatctgta 218  
DB 6122 TATGATGATGATTTTATACATTTTATACATGATTTTATATATTAGTATATATCTTAA 6181  
OY 219 aatcttccaaggaattcttctttaaataatgctgattataaagacatttgat 278  
DB 6182 AAGAT 6241  
OY 279 gtttaagtgatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 338  
DB 6242 ATTTAT 6301  
OY 339 aagataaataaagattcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 397  
DB 6302 TTTAT 6361  
OY 398 aatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 457  
DB 6362 AAT 6421  
OY 458 aagtgatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 517  
DB 6422 AT 6481  
OY 518 gttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 552  
DB 6482 ATGCTAT 6516

RESULT 14  
US-08-883-795A-36/C  
Sequence 36, Application US/08883795A  
Patent No. 5985607  
GENERAL INFORMATION:  
APPLICANT: Delcuve, Genevieve  
APPLICANT: Awang, Gregor  
TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERESKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/883,795A  
FILING DATE: 27-JUN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gravelle, Michelle  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 7841-062  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 665 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: Rh 32  
US-08-883-795A-36  
Query Match 3.3%; Score 54.4; DB 2; Length 665;  
Best Local Similarity 45.5%; Pred. No. 0.0093;  
Matches 193; Conservative 0; Mismatches 231; Indels 0; Gaps 0;  
OY 33 tttcttcttcaaggaattcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 92  
DB 454 TGTAT 395  
OY 93 aagataatctgattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 152  
DB 394 TTTAT 335  
OY 153 ttgggtcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 212  
DB 334 TTTAT 275  
OY 213 tatgtaatcttcttcaaggaattcttcttcttcttcttcttcttcttcttcttcttcttctt 272  
DB 274 TTTAT 215  
OY 273 ttgagtgtaagtgcaatcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 332  
DB 214 TTTAT 155  
OY 333 ccttgtaagataaataaagattcttcttcttcttcttcttcttcttcttcttcttcttcttctt 392





